Alternative Splicing in Plant Defence

Lesley Ann Foster

Doctor of Philosophy

Systems Biology Doctoral Training Centre

University of Warwick

September 2016

Table of Contents

List of Figures

List of Tables

Acknowledgements

First and foremost, I wish to thank my supervisor, professor Katherine Denby, for encouraging my research and for allowing me to grow as a research scientist. I am immensely grateful for her patience, motivation, and immense knowledge, without which this thesis would not have been possible.

Besides my supervisor, I would like to thank the rest of my thesis committee: Dr Miriam Gifford, Dr Sascha Ott and Dr Vardis Ntoukakis, for their insightful comments and encouragement, but also for the hard question which inspired me to widen my research from various perspectives. In addition, I would like to thank Dr Clair Hill and Dr Justyna Prusinska, without their support and guidance it would not be possible to conduct this research.

Last but not the least, I would like to thank my family: my husband Mark Everitt and to my two children Alexander and Sophie Everitt for their patient, support and continuous encouragement.

Declaration

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree.

The work presented (including data generated and data analysis) was carried out by the author except in the cases outlined below:
RNA sample preparation for RNA-Seq that was carried out by Jeanette Selby, Senior Genomics Technician, School of Life Sciences, The University of Warwick, and,

Sequencing of RNA-Seq reads that was carried out at Wellcome Trust Centre for Human Genetics

Abstract

With an increasing world population and finite resources for growing crops, producing plants that can adapt well to adverse conditions is crucial to enable adequate food production. The development of crops that have enhanced disease resistance form a core part of this strategy. We investigate the potential of Alternative splicing (AS), the mechanism whereby the same pre-mRNA can lead to different mRNA transcripts, as a mechanism for plants to regulate their defence.

We show that *Botrytis cinerea* infection has wide ranging effects on the Arabidopsis transcriptome, with approximately one third of intron-containing genes displaying *B. cinerea*-mediated differential alternative splicing (DAS), including a wide variety of defence-related genes from all stages of the defence pathway. Our work suggests, that in-line with known AS mechanisms, *B. cinerea*-mediated DAS influences the plant defence response to *B. cinerea* by enabling the plant to mount a rapid response to pathogens via circumventing the time required for transcript initiation and mRNA accumulation, altering signal transduction by affecting functional domains of proteins, as well as altering regulatory feedback loops.

For one gene, encoding a leucine-rich repeat receptor-like kinase, we demonstrate that *B. cinerea*-mediated DAS occurs, and that expression of this gene influences resistance to *B. cinerea*. We determine that DAS of this gene is altered in loss-of-function mutants of one component of the MOS4 associated complex (MAC), *AtCDC5*, which also has a defence phenotype to *B. cinerea*. With genome-wide investigations of MAC loss-of-function mutants indicating that splicing factor associated genes form highly interconnected networks. We add substantially to the body of evidence showing genome-wide DAS occurs in response to stress. We identify some putative DAS stress core genes, and splicing regulatory element motifs, which with additional work, could be used to help design crops with added disease resistance, contributing towards food security in the 21st century.

List of Abbreviations

A3SS	Alternative 3′ splice site
A5SS	Alternative 5′ splice site
ABA	Abscisic acid
AFE	Alternative first exon
AS	Alternative splicing
BH	Benjamini and Hochberg
bp	base pairs
CAMTA	Complete Arabidopsis transcriptome microarray
cDNA	complementary DNA
CEBiP	Chitin Elicitor Binding Protein
ChIP	Chromatin immunoprecipitation
Col0	Columbia 0
CSI	Constitutively spliced intron
Cq	Quantification cycle
DAMP	Damage-associated molecular patterns
DAS	Differential alternative splicing
DASed	Differentially alternatively spliced
DE	Differentially expressed
DEG	Differentially expressed gene
DEI	Differentially expressed isoforms
DGE	Differential gene expression
dpi	days post infection
EJC	Exon junction complex
ERF	Ethylene response factor
eRF3	eukaryote release factor 3
ES	Exon skipping
ESEs	Exonic splicing enhancers
ESSs	Exonic splicing silencers
ET	Ethylene

ETI	Effector-triggered immunity
FDR	False discovery rate
FP	Forward primer
FPKM	Fragments per kilobase per million mapped
GLM	Generalised linear model
GO	Gene Ontology
hnRNP	heterogeneous nuclear ribonucleoprotein
hpi	hours post infection
HR	Hypersensitive response
IGS	Indole glucosinolate
ISEs	Intronic splicing enhancers
JA	Jasmonic acid
JA-Ile	biologically active jasmonate
JAZ	Jasmonate Zim-domain
KO	Knock-out
LB	Left border of T-DNA
LRR-RLKs	Leucine rich repeat receptor like kinase
MAC	MOS4-associated complex
MAMP	Microbial-associated molecular patterns
MAPKs	Mitogen activated protein kinases
MXE	Mutually exclusive exon
ms	Mass spectrometry
NASC	Nottingham Arabidopsis Stock Centre
NCBI	National Center for Biotechnology Information
NMD	Nonsense-mediated decay
uORF	Upstream open reading frame
NSD	No significant difference
nt	nucleotide
NTC	NineTeen Complex
OGs	Oligogalacturonides

PABP	Poly(A) binding protein
PAMP	Pathogen-associated molecular patterns
PCR	Polymerase chain reaction
PG	Polygalacturonase
PIP-Seq	Protein Interaction Profile Sequencing
PRR	Pattern recognition receptor
pre-mRNA	precursor mRNA
PSI	Percentage spliced in
PTC	Premature termination codon
PTI	Pattern-triggered immunity
R genes	disease resistance genes
RBP	RNA-binding protein
RI	Retained intron
RIP-Seq	RNA immunoprecipitation coupled to sequencing
RLK	Receptor like kinase
RLP	Receptor like protein
RNA-Seq	RNA-Sequencing
ROI	Reactive oxygen intermediates
ROS	Reactive oxygen species
RP	Reverse primer
RQ-RT PCR	Real-time quantitative reverse transcription polymerase chain reaction
RRM	RNA recognition motif
RT-PCR	reverse transcription polymerase chain reaction
SA	Salicylic acid
SD	Standard deviation
SE	Skipped exon
SEM	Standard error of the mean
SF1	splicing factor 1
SJ	Splice junction

snRNPs	small nuclear ribonucleoprotein particles
SR proteins	Serine arginine rich proteins
SRE	Splicing regulatory elements
TAP	Tandem affinity purification
TF	Transcription Factor
TPM	Transcripts per million
TTSS	Type III secretion system
U2AF	U2 auxiliary factor
UTR	Untranslated region
Y2H	Yeast two hybrid

Chapter 1: Introduction

1.1 Motivation

One of the greatest challenges of the 21st century is how to feed the increasing world population. By 2050 it is predicted that the world population will increase by 2.3 billion people and that food production will need to increase by 70%(FAO, 2009). At the moment an estimated 70% of the world's water and 38% of ice-free land is used for agriculture (Holmes *et al*., 2013); to achieve this 70% increase in food production we need to farm smarter not harder.

The main pathogen focused on in this study is *Botrytis cinerea*, a fungal pathogen voted to be one of the top ten economic and commercially important pathogens (Dean *et al*., 2012). It causes serious losses in more than 200 crop species worldwide (FAO, 2009; Williamson, Tudzynski, Tudzynski, & van Kan, 2007) and is estimated to cost the food industry in excess of €1,000,000,000 a year (Dean *et al*., 2012; Holmes *et al*., 2013). With increased resistance to antifungals and infection occurring even in well controlled storage environments (Dean *et al*., 2012; Spadaro, Garibaldi, & Gullino, 2004), greater understanding of the mechanisms of plant defence against this pathogen are important to ensure food security and prevent economic losses.

We know that the plant defence response involves a large-scale change in gene expression leading to changes in metabolic and physiological processes. This transcriptional reprogramming is established as an essential control mechanism in plant defence (Birkenbihl & Somssich, 2011). A recent in depth study has shown that Arabidopsis leaf tissue undergoes genome-wide transcriptional reprogramming following infection with the fungal pathogen *B. cinerea*, detecting expression changes in approximately one third of the Arabidopsis genome during the first 48 hours post infection. It identified key transcription factors and gene clusters associated with transcriptional reprogramming in response to pathogens (Windram

et al., 2012). Now another level of complexity and control in the regulation of plant defence is beginning to emerge - Alternative Splicing (AS). Understanding how these different mechanisms ultimately regulate the proteome is important to understand how plants react to pathogens. Greater understanding of how splicing influences disease resistance could contribute towards identifying plants and/or genes for breeding and crop engineering programs to help maintain food security in the 21st century.

1.2 Splicing

RNA splicing, the post-transcriptional process where the introns of pre-mRNA are removed and exons joined together, is a conserved mechanism controlled by the spliceosome (Keren *et al*., 2010). Spliceosomes are large structures made up of five small nuclear ribonucleoprotein particles (snRNPs), each of which are made up of U-rich small nuclear RNAs (snRNAs) and a unique set of proteins, plus a variety of larger auxiliary proteins that together recognise splice sites and catalyse the two tran-esterification steps that result in splicing of pre-mRNA (M. Chen & Manley, 2009). The compositional, structural and dynamical nature of the splicing machinery enables recognition of reactive splice sites in the pre-mRNA, whilst maintaining the flexibility required for the choice of splice sites during alternative splicing (AS) (Wahl *et al*., 2009). Different proteins are known to have an effect on the spliceosome's ability to recognise splice sites adding another layer of control and complexity (Stauffer *et al*., 2010).

1.2.1 The Spliceosome

The spliceosome is responsible for both constitutive splicing, where the same splice sites are always utilised resulting in one gene producing one protein, and in AS where multiple combinations of splice sites can be utilised resulting in one gene producing multiple mature mRNAs from a single gene (Reddy *et al*., 2012a).

The spliceosome is a very dynamic large multicomponent complex that associates with a variety of protein complexes to bring about the splicing of the pre-mRNA in

the nucleus. There are two types of spliceosome that have been identified in higher eukaryotes including plants, the more abundant, major spliceosome (also referred to as the U2 type because it splices U2-dependent introns) found in all eukaryotes, and the minor (U12 type) spliceosome, which is relatively rare (0.17% of introns in Arabidopsis are U12-dependent introns and thus spliced by the minor spliceosome)(Sheth *et al.*, 2006).

In humans the major spliceosome is composed of five snRNPs and approximately two hundred proteins that associate in a stepwise, ordered manner with the pre-mRNA to bring about targeted tightly controlled splicing (Figure 1).

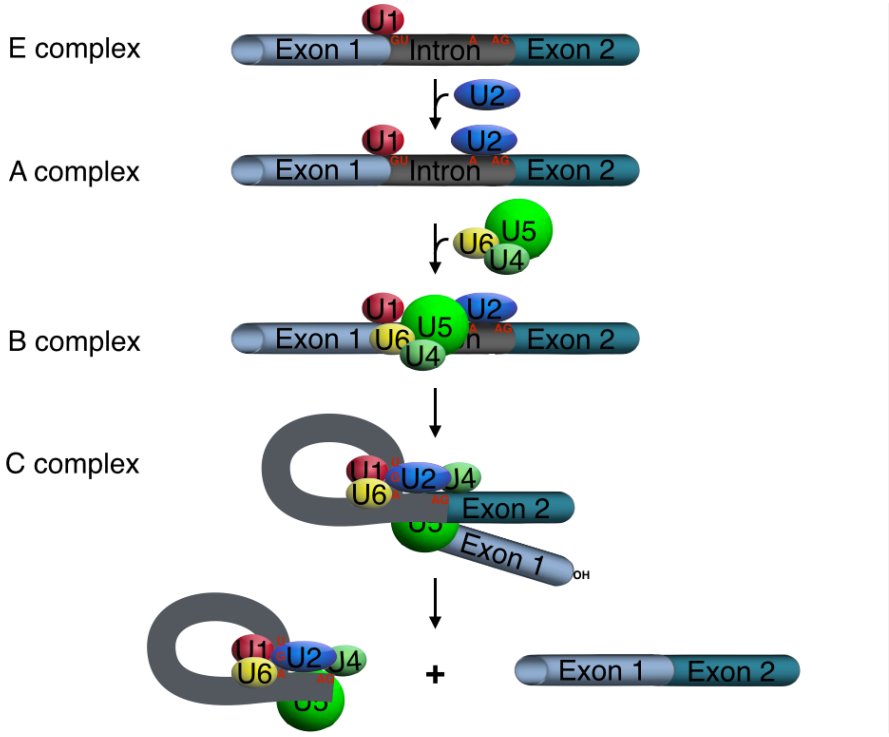


Figure 1. The major spliceosome assembles on pre-mRNA in a series of consecutive steps that produce complexes E, A, B and C. Splicing is initiated by the U1 snRNP binding to the 5' splice site. The A complex is formed when the U2 snRNP binds to the branch point. The tri-snRNP U4/U6.U5 then binds forming the B complex. A series of conformational changes results in the catalytically active C complex being formed. In an ATP-dependent manner, the mRNA is released from the spliceosome with the intron remaining in the post-splicing complex. This then dissociates, the intron is degraded and the snRNPs are reassembled and enter a new round of spliceosome assembly.

The initial step in splicing is the recognition of the 5' splice site by the U1 snRNP. In conjunction with the binding of other non-snRNP associated factors such as splicing factor 1 (SF1), which binds to the branch point, and the U2 auxiliary factor (U2AF) heterodimer, which binds to the polypyrimidine tract and 3' terminal AG, the Early (E) complex is formed (U2AF)(Berglund *et al*.,1997; Nelson & Green, 1989; Zamore & Green, 1989). This is followed by the formation of the A complex when U2 snRNP replaces SF1 at the branch point in an ATP dependent manner, with Prp5 forming a bridge between the U1 and U2 snRNP (Xu *et al*.,2004). The pre-catalytic B complex is formed when the tri-snRNP U4/U6.U5 binds. The activation of the B complex requires extensive remodelling of the RNA and protein interactions. Key events in this activation are base pairing of the U6snRNA with U2snRNA and the 5'ss, destabilisation of U1 and U4 snRNPs and the association of the Prp19/CDC5L complex in. In plants a homologue of the Prp19 complex, the MOS4 associated complex (MAC) is thought to play the same role (C. Koncz *et al*., 2012; Monaghan *et al*., 2009; Woloshen *et al*.,2011). A series of conformational changes occurs resulting in the catalytic active C complex. In an ATP-dependent manner the mRNA is released from the spliceosome. The remaining complex containing the excised intron then dissociates, the intron is degraded and the snRNPs are recycled and enter a new round of spliceosome assembly. Although spliceosomal assembly was determined in the mammalian system it is thought to be applicable to plants with sequence homology of all spliceosomal snRNA and most of the protein components of the snRNPs being conserved in plants (Lorković *et al*.,2000).

The overall assembly and splicing of the minor spliceosome is analogous to that of the major spliceosome with one exception, in contrast to the step-wise binding of the U1 followed by U2 snRNPs in the formation of the major spliceosome, U11 and U12 form a stable di-snRNP U11/U12 complex that recognises the 5' splice site and branch point site during formation of the minor spliceosome(Frilander & Steitz, 1999).

The major and minor spliceosomes differ in composition of snRNPs (U1, U2, U4 and U6 versus U11, U12, U4atac and U6atac for the major and minor spliceosome respectively), with the U5 snRNP being utilised by both, resulting in the functional difference of splicing introns with different consensus sequences(Table 1).

Table 1. Conserved sequences of U2 and U12 introns. The distinct donor site and branch point motifs associated with the two types of introns results in them being spliced by different spliceosomes. Y represents a pyrimidine. Adapted from (Lorković *et al*., 2000).

	5' splice site	Branch point	3' splice site
U2	GUAAGU	CURAY	UGYAG
U12	[A/G]UAUCCUUY	UCCUUAAC	YYCA[C/G]

1.2.2 Mechanisms and regulation of splicing

The regulation of splicing is a complex multicomponent process critical to the function of eukaryotic cells; including but not limited to cis-acting splicing regulatory elements (SREs) which bind to trans-acting factors such as serine and arginine rich (SR) proteins and the heterogeneous nuclear ribonucleoprotein (hnRNP)-like proteins, either activating or suppressing splice site recognition or spliceosome assembly, thus potentially effecting splice site choice.

1.2.2.1 Cis-acting splicing regulatory elements

SREs are short nucleotide sequences usually 6-10 nucleotides long (Reddy, Rogers, Richardson, Hamilton, & Ben-Hur, 2012b). They can be divided into four categories; exonic splicing enhancers (ESEs), exonic splicing silencers (ESSs), intronic splicing enhancers (ISEs), and intronic splicing silencers (ISSs) depending on their location in the gene and their effect; they are found to be important in both constitutive splicing and AS in animal studies (Black, 2003). Although SREs are computationally predicted to be important in both constitutive splicing and AS in plants our knowledge is limited to a few experimentally determined exonic splicing enhancers ESE motifs (Reddy *et al*., 2012b).

One of the first SREs experimentally determined in plants was a putative cis-acting element that modulates tissue-specific AS of a gene encoding chloroplast-specific ascorbate peroxidase in tobacco, comprising of a highly conserved AU-rich sequence (GU[G|C|A]UUGC[C|U]UAUUUGAAUUGCAG) that occurred over an exon/intron boundary (Yoshimura *et al*., 2002). One of the main differences between plant introns compared to yeast and metazoan introns is the high AU or U composition of plant introns (Goodall & Filipowicz, 1989). This has led some scientists to propose that the regulatory mechanisms of splicing may differ between plant and animal species, with AU- or U-rich motifs in introns acting as SREs in plant species (Yoshimura *et al*., 2002). Further evidence of this is that very few animal introns are effectively spliced in plants, and all of those that are contain AU rich regions (J. W. S. Brown & Simpson, 1998).

Pertea *et al* (2007) used computational methods to predict 84 putative ESE hexamers in the flanking regions of exons of Arabidopsis, of these 35 occurred in 9-mers shown experimentally to be exonic splicing enhancers. Interestingly the ESE hexamer motif with the highest computational score (GAAGAA) is known to function as an ESE in humans as part of the 9-mer GAAGAAGAA (Pertea *et al*., 2007; Staffa & Cochrane, 1995), indicating that plants and animals may share some SREs. At present with only a few experimentally confirmed plant SRE motifs it is impossible to say if they are conserved between metazoans and plant species or if there are distinctly different plant SREs. One important thing to note is that the work on SRE in metazoans has primarily been focused on those motifs found within exons due to the fact that skipped exons are the predominate form of AS in metazoans; in plants the most common event is retained introns, indicating that motifs found within plant intronic regions may be more important for plant species.

The first intronic cis-element, involved in AS of a plant SR gene (*SCL33*), was experimentally identified by Thomas *et al* (2012) using a GFP-intron-GFP splicing reporter in protoplasts. All known splice variants of *SCL33* were produced from the

GFP-intron-GFP construct indicating that all necessary information for proper splicing is located within the intron. Using mutational studies, Thomas *et al* (2012) found that four closely spaced GAAG repeats are important for the splicing of the *SCL33* gene with the number of repeats effecting the isoforms produced, and it functions as the binding site for the SCL33 protein in Arabidopsis. Three copies of the four-nucleotides purine unit GAAG has been identified as a putative intronic splicing regulatory motif located in intron 11 of the human dystrophin gene, resulting in an alternative exon event (Gualandi *et al*., 2003). This GAAG motif is also found as part of the experimentally verified 9-mers found to be ESEs in plants by Pertea *et al* 2007 (Pertea *et al*., 2007), indicating that the same motif may potentially act as a splicing enhancer for either an exon or intron depending on its location.

Another purine rich motif, a GAG repetitive motif, was found in the exonic regions flanking introns that are retained after undergoing heat shock in the moss *Physcomitrella patens*, potentially indicating that this motif may function as a regulatory *cis*-element in heat-mediated AS regulation (C.-Y. Chang, Lin, & Tu, 2014). In mammals it has been found that purine rich repetitive motifs can act as SRE (Ramchatesingh *et al.*, 1995; Yeakley *et al.*, 1996). Yeakely *et al* 1996 found that a repetitive GAG motif acted as a binding site the SR protein SRp40, once again indicating that there may be common elements to the splicing code of both plants and animals. Further work to identify the *trans*-acting factors associated with the *cis*-acting motifs from plants mentioned above is required to fully understand their importance and role in splicing regulation.

1.2.2.2 Trans-acting factors

The two well-known *trans*-acting factor protein classes (SR and hnRNP proteins) are conserved between plants and humans; they have a regulatory role in AS, often in an antagonistic manner (M. Chen & Manley, 2009; Reddy, 2007; Simpson *et al*., 2010). SR proteins are known to facilitate splice site recognition by binding to ESEs and can act as a general splicing factor or specific splicing modulators in a range of organisms from plants to humans (M. Chen & Manley, 2009; Lopato *et al*., 1999; Reddy, 2007).

HnRNP-like proteins repress splice site recognition by binding to splicing silencers in a variety of organisms, they are thought to play a similar role in plants (Blanchette & Chabot, 1999; M. Chen & Manley, 2009; B.-B. Wang & Brendel, 2004b).

1.2.2.2.1 SR proteins

SR proteins are a conserved family of pre-mRNA splicing factors that are concerned with the execution of constitutive splicing and alternative splice site choice in plants and animals. They can modulate AS by effecting splice site choice in a concentration and phosphorylation-dependent manner, potentially regulating tissue-specific, developmental-regulated and stress responsive AS in plants and animals (Duque, 2011).

SR proteins consist of a N-terminal domain which contains one or two RNA recognition motifs (RRMs) which bind specific RNA sequences (the SRE), and a downstream C-terminal region which contains a sequence greater than 50 amino acids long that is rich in serine and arginine (the RS domain) that recruits other proteins or in some cases contacts the pre-mRNA branch point; they can also contain signals for cellular localization (Caceres, Screaton, & Krainer, 1998; J. C. Long & Caceres, 2009; H. Shen & Green, 2006). According to standard nomenclature in mammals the RS domain has 40% RS content with consecutive RS or SR repeats, whereas in plants the RS domain has a RS content of 20% RS or SR dipeptides (Barta, Kalyna, & Reddy, 2010; Manley & Krainer, 2010). Barta *et al* 2010 have proposed that in Arabidopsis the SR proteins can be divided into six sub families (Figure 2). The SR subfamily has four members and consists of two RRM domains, the second of which contains the evolutionary conserved SWQDLKD motif followed by a SR domain featuring characteristic SR dipeptides; the RSZ subfamily contains one Zinc knuckle and comprises three members; the SC subfamily consisting of a single RRM domain followed by a SR domain (one member in Arabidopsis); the SCL subfamily which is very similar to the SC subfamily but contains a N-terminal charged extension (four members in Arabidopsis); the RS2Z subfamily which contains two Zinc knuckles with an additional SP-rich region following the RS domain (two members in Arabidopsis)

and the RS subfamily that consists of two RRM domains followed by a RS domain (four Arabidopsis members) (Barta *et al*., 2010). Using this nomenclature two previously classified SR proteins SR45 and SR45a are no longer classed as SR proteins, which is consistent with their mammalian homologues that have been declassified as SR proteins by Manley and Krainer (Barta *et al*., 2010; Manley & Krainer, 2010).

Subfamily name	Protein/gene name	Aliases	Accession	Human homologues
SR Subfamily 	At-SR30	At-SRp30	At1g09140	SRSf1/SF2/ASF
	At-SR34	At-SRp34, SR1	At1g02840	
	At-SR34a	At-SRp34a	At3g49430	
	At-SR34b	At-SRp34b	At4g02430	
RSZ subfamily 	At-RSZ21	At-RSzp21, SRZ21	At1g23860	SRSF7/9G8
	At-RSZ22	At-RSzp22, SRZ22	At4g31580	
	At-RSZ22a	At-RSzp22a	At2g24590	
SC subfamily 	At-SC35	At-SC35	At5g64200	SRSF2/SC35
SCL subfamily 	At-SCL28	At-SCL28	At5g18810	Plant specific class
	At-SCL30	At-SCL30	At3g55460	
	At-SCL30a	At-SCL30a	At3g13570	
	At-SCL33	At-SCL33, SR33	At1g55310	
RS2Z subfamily 	At-RSZ32	At-RSZ32	At3g53500	Plant specific class
	At-RSZ33	At-RSZ33	At2g37340	
RS subfamily 	At-RS31a	At-RSp31a	At2g46610	Plant specific class
	At-RS31	At-RSp31	At3g61860	
	At-RS40	At-RSp40, At-RSp35	At4g25500	
	At-RS41	At-RSp41	At5g52040	

Figure 2. Domain Architecture of the Arabidopsis SR Protein Subfamilies. The SR subfamily proteins contain an evolutionary conserved SWQDLKD motif in their second RRM with a downstream RS domain featuring characteristic SR dipeptides. The RSZ subfamily consists of SR proteins that contain one Zinc knuckle. The SC subfamily proteins have a single RRM followed by a downstream RS domain. The plant-specific SCL subfamily (SC35-like) proteins are similar to the SC subfamily but have an N-terminal charged extension. The proteins of the plant-specific RS2Z subfamily possess two Zinc knuckles and have an additional SP-rich region following the RS domain. The plant-specific RS subfamily proteins contain two RRMs (without the SWQDLKD motif) followed by the RS domain rich in RS dipeptides. Red oval represent RRM domain, purple ovals represent RRM domain with the SWQDLKD motif, blue rectangles represent RS domain consisting of SR dipeptides, yellow circles represent Zinc knuckles, turquoise rectangles represent RS domain, purple rectangle represent N-terminal charged extension and green rectangles represent a SP-rich region; adapted from Barta *et al* 2010.

Flowering plants contain the largest number of SR proteins amongst eukaryotes, 18 in Arabidopsis (Barta *et al*., 2010), compared to seven in Caenorhabditis elegans (Longman, Johnstone, & Caceres, 2000), and 12 in humans (Manley & Krainer, 2010). This is thought to be due to genome amplification, particularly inter-chromosomal duplication events, with at least 12 Arabidopsis genes that encode SR proteins being located on duplicated regions. Most of the duplicated genes display different spatiotemporal expression patterns indicating functional diversification (Kalyna & Barta, 2004).

SR proteins interact with RNA through the RRM domain, providing binding specificity, the RS domain aids in spliceosomal assembly via binding of other spliceosomal proteins (Reddy *et al*., 2013). In mammalian cells during early spliceosome assembly the binding of U1 snRNP to the 5'SS is mediated by the SR protein SRSF1, via direct interaction with the RS motif of the U1-70K component of the U1 snRNP in a phosphorylation-dependent manner (Cho *et al*., 2011). This SR protein along with SC35 has been implicated in the bridging of the 5' and 3' splice sites via interactions with U1-70K and the 35Kd subunit of the splicing factor U2AF ($U2AF^{35}$) (J. Y. Wu & Maniatis, 1993). SR proteins have also been implicated in the incorporation of the tri snRNP complex (U4/U6.U5 snRNP) into the spliceosome (Roscigno & Garcia-Blanco, 1995). It is thought that similar interactions between SR proteins, RNA and spliceosomal components occur in plant systems. Lorković *et al* 2004 showed in immune-precipitation experiments the Arabidopsis homologue of SRSF1 (AtSR34) and AT-U1-70K (the Arabidopsis version of the U1-70K component of the U1 snRNP) protein co-expressed in protoplast showing a clear interaction (Lorković, Lopato, Pexa, Lehner, & Barta, 2004). However Golovkin and Reddy *et al* 1999 showed that four other Arabidopsis SR proteins (At-SCL33, At-SRZ21, At-SRZ22, and the SR-like protein SR45) interact with the Arabidopsis U1 snRNP 70-kDa protein, and a Clk/Sty protein kinase (AFC-2) from Arabidopsis phosphorylated all four of these SR proteins (Golovkin & Reddy, 1999). This indicates that the mechanisms may be similar between plants and animals, but the specifics of early

stages of spliceosome assembly, and intron recognition in plants is likely to be different.

1.2.2.2.2 hnRNP proteins

Traditionally SR proteins were thought of as activators of splicing with hnRNP proteins acting as repressors by binding to the pre-mRNA preventing other splicing factors from accessing the binding site (B.-B. Wang & Brendel, 2004b). However it is now thought that the type of action may be more context dependent (Wachter *et al*., 2012). For example, in mice, the hnRNP-like proteins belonging to the neuro-oncological neural antigen (NOVA) family recognise an YCAY motif that can act as either an ESS if it is in an exon preceding an AS exon or as an ISS if it is in an intron following an AS exon (Ule *et al*., 2006).

The RNA regulatory map appears to be highly conserved between insects and mammals, for example the *Drosophila melanogaster* ortholog of mammalian NOVA1 and NOVA2, Pasilla (PS), has regions enriched for YCAY repeats upstream and within PS-repressed exons and downstream from PS-activated exons, consistent with the location of these motifs near NOVA-regulated exons in mammals. The target gene orthologs regulated by PS and NOVA1/2 are almost entirely non-overlapping potentially indicating that the regulatory codes of individual RNA-binding proteins may be conserved between species but the regulatory modules controlled by these proteins are evolving (Brooks *et al*., 2011).

The Polypyrimidine tract-binding (PTB) proteins are probably the most intensively studied hnRNP proteins, with homologues found between the mammalian system and plants. In HeLa cells, Lin and Patton *et al* 1995 showed that PTB effects the splicing of α-tropomysin (α-TM) (C. H. Lin & Patton, 1995). The second and third exon of α-TM are mutually exclusive and PTB inhibits the splicing that joins the first and third exon (1-3 splicing) with U2AF antagonising this splicing. In contrast, SR proteins activate the splicing that joins the first and second exon (1-2 splicing) demonstrating how hnRNP and SR proteins can work together to produce AS

isoforms (C. H. Lin & Patton, 1995). In Arabidopsis there are three homologues of the human PTB (AtPTB1, AtPTB2 and AtPTB3) (B.-B. Wang & Brendel, 2004b). Simpson *et al* 2014 showed using a protoplast transient expression system that the Arabidopsis proteins AtPTB1 and AtPTB2 reduced the inclusion/splicing of a novel potato invertase mini-exon splicing reporter, indicating that these proteins can repress plant intron splicing; the co-expression of NpU2AF^{65} alleviated this repression, indicating that PTB and U2AF also compete for pre-mRNA binding in plants (Simpson *et al*., 2014). This indicates the mechanisms of activation and repression of splice sites by hnRNP may be conserved between mammals and plants.

Plants are thought to contain more hnRNP proteins than humans; Wang and Brendel *et al* 2004 computationally identified 35 other potential hnRNP proteins in Arabidopsis (not including the PTB hnRNP proteins) that have human homologues and 17 plant-specific potential plant hnRNP proteins, compared to 20 hnRNPs found in humans (identified as hnRNP A to hnRNP U) (B.-B. Wang & Brendel, 2004b).

1.2.2.3 The MOS4 associated complex (MAC)

The MAC is an evolutionary conserved spliceosome-associated nuclear complex consisting of 24 proteins, MOS4 and 23 proteins that interact with it, including ATCDC5 which is shown to interact with PRL1 and MAC3 *in planta* (Figure 3) (Monaghan *et al*., 2009; Palma *et al*., 2007).

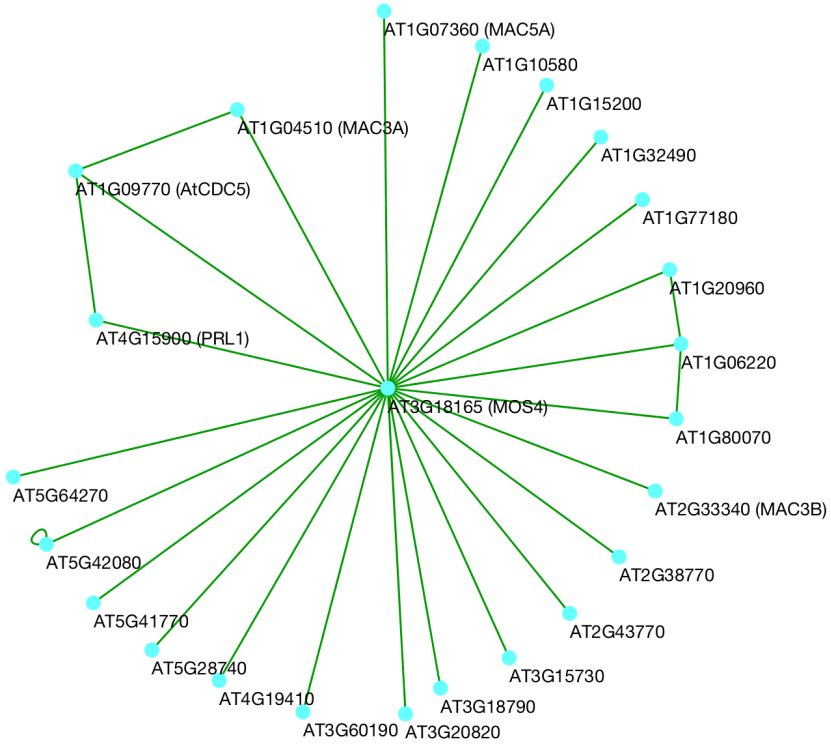


Figure 3. Network view of the MAC. The MAC consists of MOS4 and 23 proteins that interact with it. AtCDC5 has been shown to interact with PRL1 and MAC3A in planta. Network constructed in cytoscape (Shannon *et al*., 2003), utilising immune-precipitation evidence from Palma *et al* (2007) and Monaghan *et al* (2009).

The 24 components of the MAC were identified by Monaghan *et al* (2009) via immuno-affinity purification of the MAC using *mos4* knock-out (KO) transgenic lines (*mos4-1*) complemented with MOS4-HA, a full-length MOS4 containing a C- terminal triple hemagglutinin (HA) epitope tag under the control of the native promoter of mos4, followed by protein sequencing using mass spectrometry (MS). Of these 24 proteins, 19 share homology to proteins found in the NineTeen Complex (NTC) in yeast and the PRP19 complex (also called the CDC5L complex) in humans (Monaghan *et al*., 2009).

The Prp19 complex was first identified in humans as a non-snRNP splicing complex. However, it has now been shown to be involved in a variety of processes including genome maintenance, protein degradation by the proteasome, transcription elongation and functions in lipid droplet biogenesis (Chanarat & Sträßer, 2013). In humans, the PRP19 complex associates with the U5 snRNP to help bring about the conformational change from the pre-catalytic spliceosomal B complex to the

activated complex C, remaining associated with the spliceosome throughout the second catalytic step (Chan & Cheng, 2005; Chan *et al*., 2003; Makarov *et al*., 2002). The MAC resembles this human complex with the core components of the MAC having homologues of proteins known to be essential for spliceosome activation; AtCDC5, MAC3, MAC5, MOS4 and PRL1 associate with the spliceosome and due to their homology with the PRP19 components are predicted to function in a similar manner, bringing about the conformational change required form the activated C complex in plants (C. Koncz *et al*., 2012; Woloshen *et al*., 2011) (Figure 4).

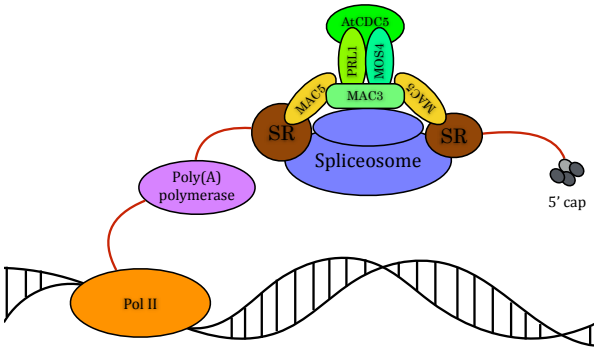


Figure 4. Core components of the MOS4-associated Complex (MAC) associate with the spliceosome. In plants the MAC components AtCDC5, MAC3, MAC5, MOS4 and PRL1 associate with the spliceosome and are thought to be involved in splicing. Adapted from Woloshen *et al* (2011).

1.3 Alternative Splicing

It has been known for many years that AS is important for gene regulation and proteome diversity in animals, but has only recently begun to be investigated in the plant kingdom (Barbazuk *et al*., 2008). There are five main types of AS events; Intron Retention (IR), Alternative 3′ Splice Site (A3SS), Alternative 5′ Splice Site (A5SS), Skipped Exons (SE), which includes alternative first exon (AFE) and alternative last exon ALE), and mutually exclusive exons (MXE) (E. T. Wang *et al*., 2008) (Figure 5).

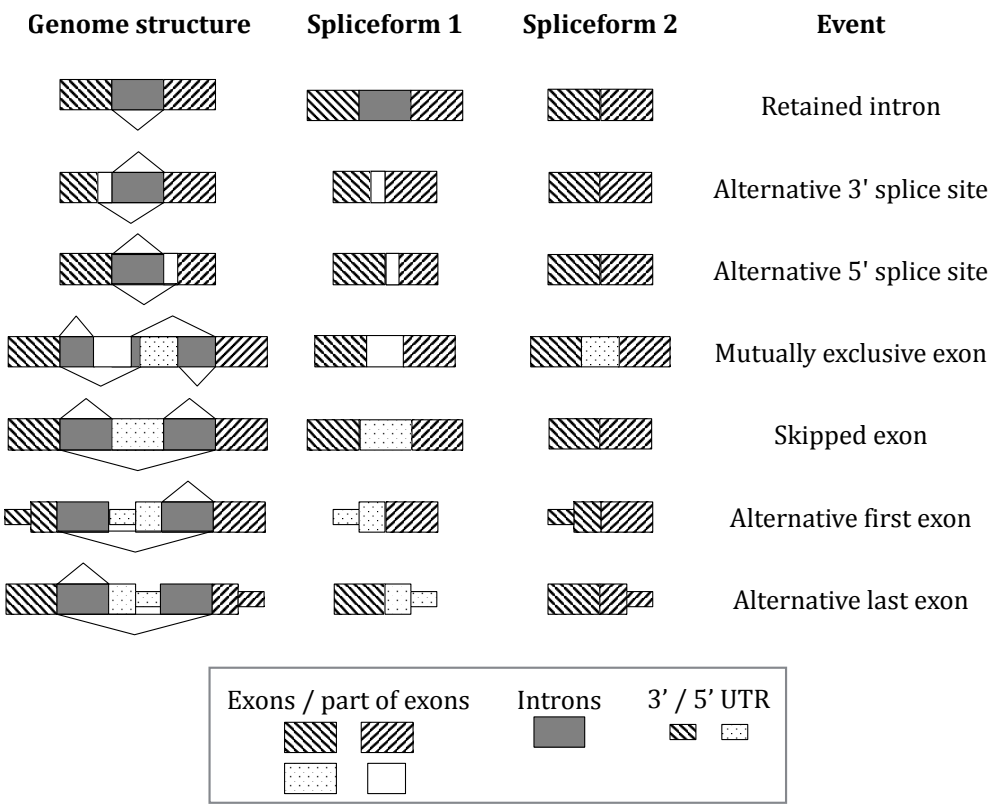


Figure 5. The main Alternative Splicing events. The first column illustrates the intron-exon structure of the AS event, columns 2-3 represent the possible spliceforms arising from such events and column four describes the type of event. Diagonal lines represent the alternative splicing events. A gene may undergo one or more of these basic events individually to produce multiple isoforms. More complex AS events may also occur when a combination of these basic events overlaps possibly due to connections in regulation or function.

These events can occur individually in a gene, or a more complex AS event may occur where multiple basic events are combined and overlap each other, possibly due to connections in regulation or function (Foissac & Sammeth, 2007). The resulting mRNA may produce multiple proteins with potentially different localisation, concentrations or domains from a single gene.

All five types of events occur in both the animal and plants kingdoms, however the proportion of each type of event differs, SE are the predominant event type in multicellular animals whereas RI events are the most numerous event type in multicellular plants (Campbell *et al*., 2006; Iida *et al*., 2009; Marquez *et al*., 2012; Ner-Gaon *et al*., 2004; Sugnet *et al*., 2004; Thanaraj *et al*., 2004; B.-B. Wang & Brendel, 2004a).

The predominance of SE in animals and RI in plants has led to the theory that splicing may predominantly occur via an exon definition model in mammals and an intron definition in plants (McGuire *et al*., 2008). Under the exon definition model, splice sites either side of an exon are recognised as one unit compared to the intron definition model where the splice sites either side of an intron are recognised as the splicing unit; it is thought that the exon definition model is utilised when introns separating exons are large (>250 bp) compared to the intron definition model which is utilised when the introns are short (<250 bp) (De Conti *et al*., 2012) (Figure 6).

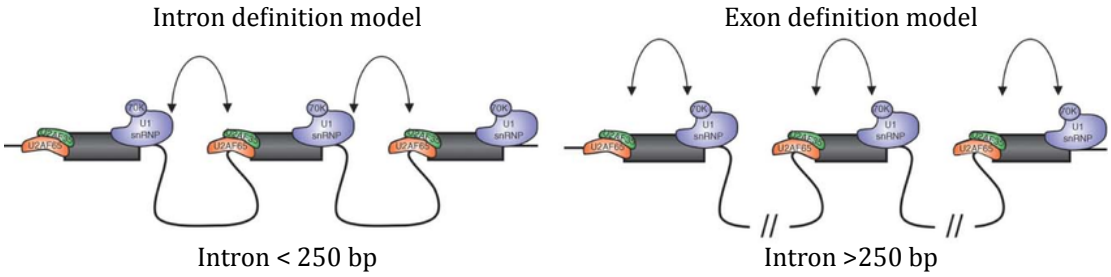


Figure 6. Intron and exon definition models. In the Intron definition model (left hand side) pairing between the splice sites occurs across an intron where long exons are separated by short (<250 bp) introns. Whereas in the exon definition model the splice sites are paired across exons. This occurs when they are separated by long (>250 bp) introns. Image reproduced from De Conti *et al* (2012).

McGuire *et al* 2008 found the number of RI and SE in plants was intermediate compared to animals, which had the highest proportion of SE events, and to fungi and most protists, which had the highest proportion of RI events (McGuire *et al*., 2008). This agrees with the theory that the model utilized correlates to intron size, because on average animals have the largest intron sizes, plants have intermediate intron sizes, and fungi have short intron lengths (Deutsch & Long, 1999). Both the intron definition and the exon definition model can occur in the same organism and experiments by Kennedy *et al* 1998 showed that in *D. melanogaster* both models can be utilized by the same mRNA. It is likely that plant species utilize both of these mechanisms because length of plant introns covers a large range and the average intron length is close to the point where switching between the intron definition model and exon definition model is thought to occur; recent publications based on

genome sequencing report an average intron length of 168 bp, (Arabidopsis Genome Initiative, 2000).

Knowledge about AS in plants has grown rapidly over the last 15 years, partly due to the advent of high throughput technologies such as RNA-Sequencing (RNA-Seq). RNA-Seq is a high throughput technology that sequences cDNA enabling comprehensive and quantitative analysis of the transcriptome; mRNA is either fragmented then reverse transcribed into cDNA, or as is in the case of Illumina RNA-Sequencing, cDNA is reversed transcribed and then fragmented. This fragmentation can occur either chemically or by using a nebulizer. Adaptor sequences are then ligated to either one or both ends of the cDNA fragments. A next-generation sequencing platform such as Roche's 454, Illumina, or Applied Biosystems SOLiD is used to generating millions of short reads. The cDNA can be sequenced from one end (single-end sequencing) or both ends (pair-end sequencing) producing reads that are 30-400 bp in length, depending on the platform used. These reads can then be aligned to either a reference genome, transcriptome or *de novo* reconstruction can be undertaken after which the transcriptional architecture investigated.

Recent RNA-Seq work carried out under multiple developmental and stress conditions indicates that AS occurs in between 44 - 61% of plant multi exon genes, with between 40 and 65% of the events being RI (Chamala *et al*.,2015; Ding *et al*., 2014; Howard *et al*., 2013; Marquez *et al*., 2012; R. Zhang *et al*., 2016). The differences in amount of AS seen in the experiments may be due the transcriptome and genome used, because how well annotated genomes and transcriptomes are has a profound effect on the quantification of isoforms, especially when analysing RNA-Seq data (R. Zhang *et al*., 2016). In addition, differences in analysis, read depth sequencing or library preparation could also account for the variation seen.

1.3.1 Alternative splicing coupled to nonsense-mediated decay.

Genes that are alternatively spliced may produce transcripts that contain premature termination codons (PTCs), potentially resulting in targeted degradation by

nonsense-mediated decay (NMD). When coupled to AS this mechanism has been implicated in the fine-tuning of the expression of important regulatory genes in Arabidopsis Kalyna *et al*., (2012).

During translation, NMD discriminates between PTC and authentic stop codons via *cis* elements that are present downstream which trigger the formation of a functional NMD complex on these mRNAs, targeting them for rapid degradation (Kerényi *et al*., 2008). The *cis* elements are thought to differ between the mammalian system and those in yeast and invertebrate systems. The first model, which is characteristic of yeast and invertebrate transcripts, is the *faux* 3'UTR model, where efficient termination and stability of mRNA is dependent on the 3'UTR being properly configured (Amrani *et al*., 2004).

In this model, a long 3'UTR inhibits the interaction between Poly(A) binding protein (PABP) and the eukaryote release factor 3 (eRF3) component of the terminating ribosome, causing translation termination to be inefficient. This enables UPF1 to bind to eRF3, UPF1 recruits UPF2 and UPF3 forming the NMD complex, resulting in the rapid degradation of the mRNA (Amrani *et al*., 2004; 2006). Tethering of PABP downstream of a PTC but in close proximity of it protects the mRNA from degradation; this is the equivalent of having a shorter UTR and thus supports the *faux* model (Amrani *et al*., 2004; Behm-Ansmant *et al.*, 2007).
The second model, which is characteristic of mammalian NMD, is the intron-based NMD mechanism (Figure 7).

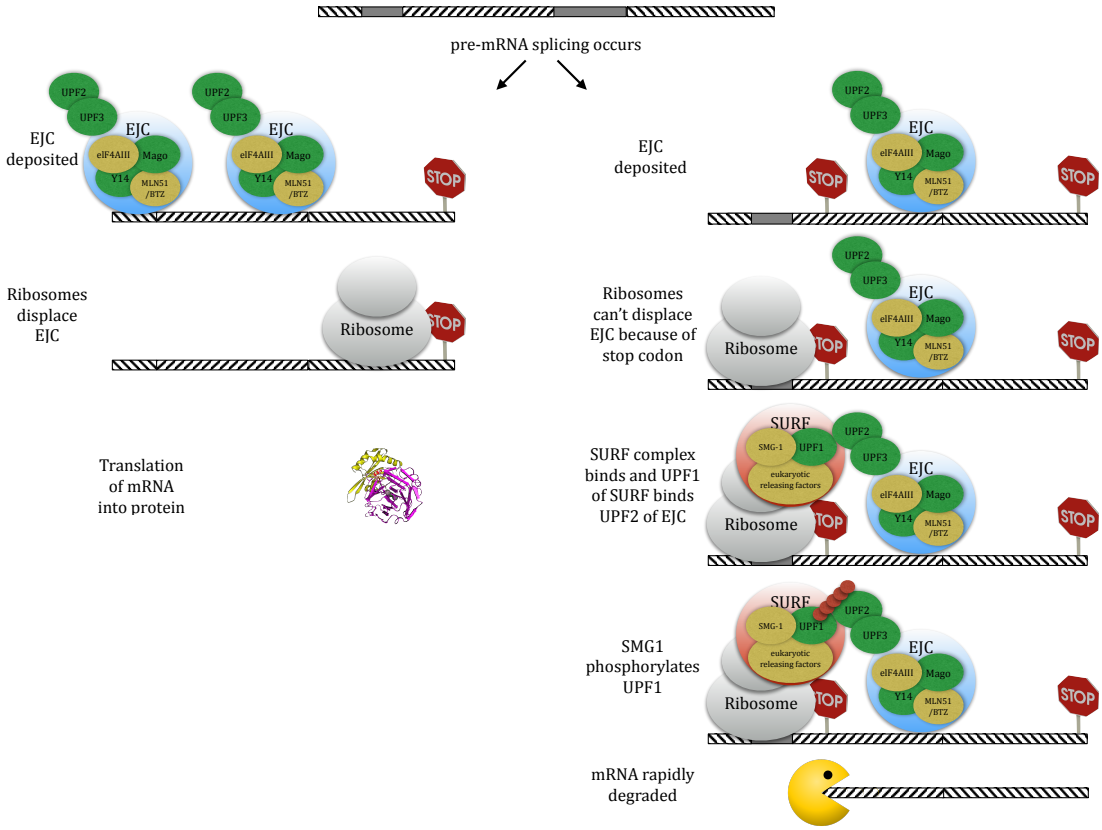


Figure 7. The intron-based NMD mechanism. Left hand panel; mRNA is translated into protein. The exon junction complex (EJC) is deposited approximately 20-25 nucleotides upstream of the exon-exon junction, ribosomes traverse the mRNA and displace any EJCs in its path enabling efficient translation of protein to occur. Right hand panel; mRNA is rapidly degraded by NMD mechanisms. The EJC is deposited and ribosomes displace any EJCs in its path, however the EJC-UPF2-UPF3 complex cannot be displaced because it is downstream of the premature stop codon. In this example because of a PTC occurring due to a retained intron event, however this may occur due to DNA architecture, for example if the stop codon is not situated on the last exon, or due to other AS event types. The SURF complex binds to the ribosome at the termination of translation and UPF1 binds to UPF2 in the EJC. This causes SMG1 to phosphorylate UPF1, resulting in recruitment of mRNA decay systems to the mRNA, which is rapidly degraded. Green indicates known plant homologues.

When an intron is spliced an exon junction complex (EJC) is deposited approximately 20-25 nucleotides (nt) upstream of the exon-exon junction on the pre-mRNA (Le Hir *et al*., 2000). The mammalian EJC, comprising of four core proteins (Y14, Mago, MLN51/BTZ and eIF4AIII), provides an anchoring point for the NMD factors UPF2 and UPF3 (Ballut *et al*., 2005; Le Hir *et al*., 2001). During translation the EJC-UPF3-UPF2 complex is displaced by the ribosomes unless it is positioned downstream of a stop codon (Y.-F. Chang et al., 2007). At the termination of translation the SURF complex

(consisting of UPF1, SMG-1 and eukaryotic releasing factors) binds to the ribosome, if the EJC-UPF2-UPF3 complex is still bound, for example if it was not displaced by the ribosome because it is downstream of the stop codon, the UPF1 component of the SURF complex binds UPF2 resulting in SMG-1 mediated phosphorylating UPF1, which in turn recruits mRNA decay systems to the mRNA (Kashima et al., 2006). Lastly SMG-5, SMG-6 and SMG-7 recruit PP2A phosphatase to dephosphorylate UPF1 (Okada-Katsuhata et al., 2012; Yamashita et al., 2005). If the EJC is located downstream of a stop codon but within close proximity to it then the ribosome can remove them, this results in mammalian transcripts only being targeted for NMD if the stop codon is >50 nt upstream of the exon-exon junction (Nagy & Maquat, 1998).

Although the intron based model is the characteristic model of NMD in mammals, in humans long 3'UTRs can also target transcripts for NMD. Kurosaki *et al* (2013) showed that there is a linear increase in the binding of UPF1 with respect to 3'UTR length in transcripts that contain a PTC, with greater binding occurring when there is an EJC present in the 3'UTR; indicating that mRNAs without a 3'UTR EJC but with an abnormally long 3'UTR can be NMD targets, however NMD is not as efficient compared to transcripts that contain a 3'UTR EJC.

In plants, long 3'UTR and 3'UTR located introns can act as NMD triggers, implying that both models occur in plants (Hori & Watanabe, 2007; Kerényi *et al*., 2008; Kertész *et al*., 2006). An agroinfiltration-based transient NMD assay system in *Nicotiana tabacum* and *Nicotiana benthamiana* leaves identified that unusually long 3'UTRs or the presence of introns in the 3'UTR can subject mRNAs to NMD (Kertész *et al*., 2006). In addition Hori *et al* (2007) used an Agrobacterium transient expression assay in *N. benthamiana* to analyse the positions of PTCs. They determined that termination codons located distant form the mRNA 3'termini (which would result in a long 3'UTR), or greater than 50nt upstream of the 3' most exon-exon junction (i.e. presence of a 3'UTR intron which when spliced results in an EJC located downstream of the stop codon) targeted transcripts for NMD.

Although the NMD pathway is not well characterised in plants, some of the core mammalian NMD machinery have plant orthologs that have been functionally linked to degradation of transcripts by NMD; in particular SMG7 and the UP FRAMESHIFT (UPF) proteins UPF1, UPF2 and UPF3 and the EJC components Mago and Y14 (Arciga-Reyes *et al*., 2006; Hori & Watanabe, 2005; Kerényi *et al*., 2008; Yoine *et al*., 2006). To determine the molecular mechanisms of plant NMD Kerenyi *et al* (2006) used a virus-induced gene silencing-NMD (VIGS-NMD) assay in *N. benthamiana* leaves; they demonstrated that the NMD proteins SMG7, UPF1, UPF2 and UPF3 are required for both types of NMD mechanisms in plants, whereas the plant homologues of the EJC components Mago and Y14 are only required for the intron-based method, this implies that the molecular mechanisms of NMD are similar across species (Kerényi *et al*., 2008).

In addition to long 3'UTR and 3'UTR located introns acting as NMD triggers, upstream open reading frames (uORF) have also been identified as potential triggers for NMD. Saul *et al* (2009) suggested that the presence of an uORF in the Arabidopsis vacuolar transporter gene, *AtMHX*, subjects the transcripts to NMD. Recently Kalyna *et al* (2012) utilised a high resolution RT- PCR (HR RT-PCR) panel of Arabidopsis NMD factor mutants to identify genome-wide endogenous Arabidopsis isoforms that increase in abundance when NMD is impaired; these experiments elucidated some novel NMD triggers in plants; most notably uORFs that overlap the main start codon act as a new trigger for NMD, and NMD is induced if 3'UTRs were >350 nt. These experiments also indicated that the type of AS event might play a role in targeting transcripts for NMD. They found that the majority of intron retention transcripts that possess NMD features such as PTCs, downstream splice junctions and long 3'UTR are insensitive to NMD; however other AS events in the same genes that generated transcripts with PTCs in similar positions were subject to NMD (Kalyna *et al*., 2012). Göhring *et al* 2014 determined that PTC transcripts generated by RI events were generally not targeted for NMD in plants because these transcripts were retained in the nucleus, and thus could not undergo NMD due to the fact that RNA degradation

via the NMD pathway occurs in the cytoplasm. Mammalian and yeast intron retention transcripts have been shown to be sensitive to NMD, indicating that plants may have a different NMD mechanism compared to other organisms (de Lima Morais & Harrison, 2010; Kalyna *et al*., 2012; Sayani *et al*., 2008).

1.3.2 Alternative splicing and trans acting factors

Both the SR and hnRNP proteins have been shown to undergo AS, with NMD coupled to AS occurring in both protein families indicating that this may be a mechanism of fine-tuning the levels of these trans acting splicing factors.

1.3.2.1 SR proteins

Expression levels of SR proteins are differentially expressed in different organisms and developmental stages indicating that this may be one mechanism for organ-specific AS in plants due to the fact that SR factors regulate alternative splicing (Palusa *et al*., 2007). With over 88.9% of SR proteins undergo AS in Arabidopsis implying that they regulate and are regulated by AS mechanisms (Chamala *et al*., 2015; Richardson *et al*., 2011). Overexpression of *atSR30* in transgenic Arabidopsis plants alters the splicing patterns of a variety of endogenous plant genes, including *atSR30* itself, as well as down-regulating the mRNA encoding the full-length atSR34/SR1 protein, an SR protein with which it shares strong sequence similarities, with both these genes are deemed to encode SF2/ASF-like proteins due to their strong sequence similarities to human SF2 (Lopato *et al*., 1999). The splicing patterns of the mRNA encoding the SF2/ASF-like proteins *atSR30* and *atSR34* are altered in transgenic Arabidopsis plants overexpressing the plant specific SR protein *atRSZ33* (Kalyna *et al*., 2003), indicating that SR proteins can regulate AS of their own genes as well as other endogenous plant genes. Further evidence that SR proteins can modulate splicing in plants comes from Isshiki *et al* (2006) work in rice protoplasts where they showed that the SR proteins RSp29 and RSZ23 enhanced splicing and altered the A5SS of the first intron of the Waxy^{b} genes; RSp29, a plant specific SR protein, and RSZ23, an SR protein homologous to human 9G8, enhanced splicing at the distal and proximal site respectively. In transgenic rice plants over expressing the

plant specific SR protein RSZ36 or the rice homolog of human SF2/ASF, SRp33b, alters the splicing patterns of their own pre-mRNAs and those of other SR proteins (Isshiki *et al*., 2006). Taken together these findings indicate that both plant specific SR proteins, and those with mammalian homologues, play a role in modulating splicing by regulating the mRNA splicing of plant genes via auto regulation and regulation of endogenous plant genes (including those that encode other SR genes proteins).

SR proteins have been shown to be regulated by AS coupled to NMD (Kalyna *et al*., 2012), in humans AS coupled to NMD of SR proteins is frequently used as an auto regulation mechanism and this may also be the case in plants. Paulsa *et al* 2010 found that in Arabidopsis mutants lacking *upf3*, one of the core components of the NMD pathway, there is a several fold increase in the transcript of SR30 (Palusa & Reddy, 2010). As mentioned above SR30 is found to be auto regulated, thus AS coupled to NMD may be one mechanism by which SR protein levels are fine-tuned. However not all SR transcripts containing PTC accumulated in the *upf3* mutants plants indicating that they may produce truncated proteins; given the fact that SR proteins have multiple functional domains these transcripts may produce protein with an altered function (Palusa & Reddy, 2010).

The splicing patterns of SR proteins are also thought to play a functional role in the plant stress responses. SR proteins expression levels are not dramatically altered in response to stress (Duque, 2011), however SR proteins are extensively DAS in response to stress, which in turn may affect the splicing of other genes (Table 2).

Table 2. SR genes known to undergo stress-mediated DAS.

Gene Identifier	Name	DAS under stress	Reference
At1g02840	SR34/SR1	Temperature, hormones, heat, and cold	(Lazar & Goodman, 2000; Palusa *et al*., 2007)
At1g09140	SR30	Salt, *Pst*, high light, and heat.	(Ding *et al*., 2014; Filichkin *et al*., 2010; Howard *et al.*, 2013; Palusa *et al*., 2007)
At1g16610	SR45	Low temperature, heat, and drought	(Gulledge *et al*., 2012; Tanabe *et al*., 2007)
At1g23860	RSZ21/SRZ-21	-	-
At1g55310	SCL33	Salt, hormones, salt, cold, heat and glucose	(Ding *et al*., 2014; Palusa *et al*., 2007)
At2g24590	RSZ22a	-	-
At2g37340	RSZ33	Salt, heat, and glucose	(Ding *et al*., 2014; Palusa *et al*., 2007)
At2g46610	RS31A	Heat	(Palusa *et al*., 2007)
At3g13570	SCL30a	Salt and heat	(Ding *et al*., 2014; Palusa *et al*., 2007)
At3g49430	SR34a	Salt	(Ding *et al*., 2014)
At3g53500	RSZ32	Salt and heat	(Ding *et al*., 2014; Palusa *et al*., 2007)
At3g55460	SCL30	Salt	(Ding *et al*., 2014)
At3g61860	RS31	Salt and cold	(Ding *et al*., 2014; Palusa *et al*., 2007)
At4g02430	SR34b	Heat, cold, hormones, and salt	(Palusa *et al*., 2007)
At4g25500	RS40/RSP35	Salt, cold and heat	(Ding *et al*., 2014; Palusa *et al*., 2007)
At4g31580	RSZ22/SRZ-22	-	-
At5g18810	SCL28	-	-
At5g52040	RS41	Salt and heat	(Ding *et al*., 2014; Palusa *et al*., 2007)
At5g64200	SC35	-	-

1.3.2.2 hnRNP

Of the *Arabidopsis* hnRNP proteins with human homologues, a superfamily of 21 glycine-rich RNA-binding proteins (RBPs) that were homologs of human hnRNP A1 and hnRNP A2/B1 was identified; this family can be divided into two subfamilies

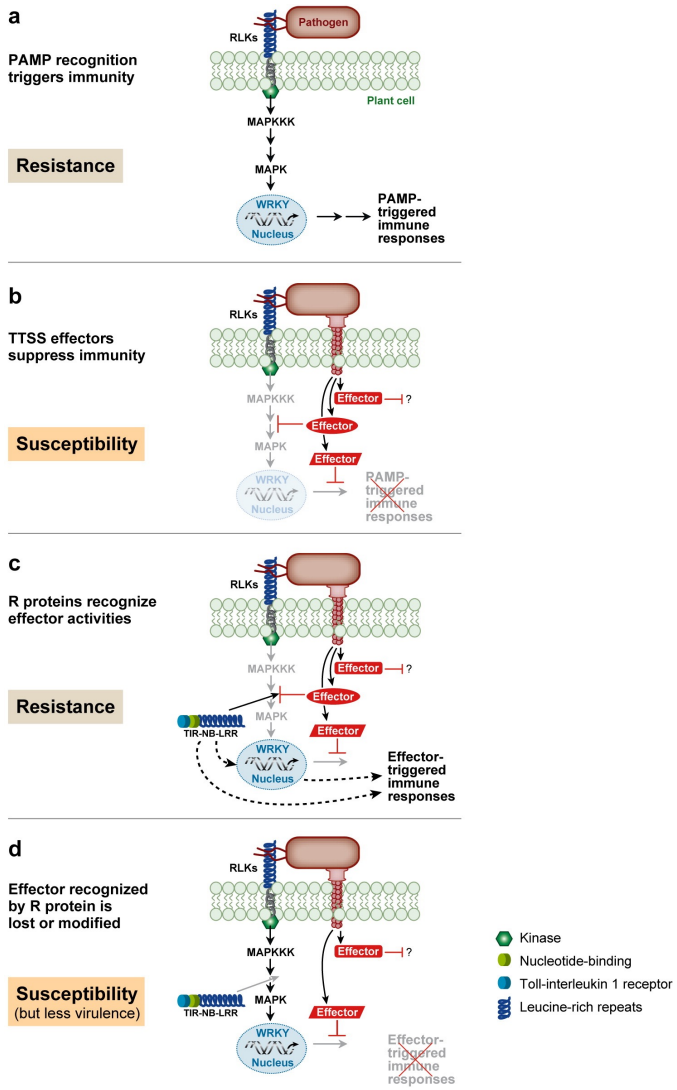
based on the number of RRM domains, 8 proteins contain one RRM domain and the remaining 13 contain two RRM domains(B.-B. Wang & Brendel, 2004b). Of these RBPs with one RRM domain AtGRP7 and its paralog AtGRP8 have been shown to regulate AS of pre-mRNA in an auto and cross regulatory manner (Schöning *et al*., 2008; Staiger *et al*., 2003). This regulation is thought to occur via AS coupled to NMD; AS transcripts of *AtGRP7* and *AtGRP8* contain PTCs, in NMD-impaired mutant plants and under cyclohexamide treatment accumulation of these mRNAs transcripts occurs indicating that those transcripts with the PTC are targets for NMD (Schöning *et al*., 2007; 2008; Staiger *et al*., 2003). Transgenic plants that have a point mutation of a conserved arginine to glutamine within the RNA recognition motif of *AtGRP7* (*AtGRP7-RQ-ox*) display a reduction in its binding affinity, which disrupts the auto-regulation of *AtGRP7* and the regulation of downstream targets (Schöning *et al*., 2007). Streitner *et al* (2012) utilized a HR RT–PCR-based AS panel to determined that 20% of analysed AS events displayed changes in the ratios of AS isoforms under ectopic *AtGRP7* expression and approximately half of these were also altered by the paralog *AtGRP8*, indicating that they may the share downstream targets. Several of the transcripts with altered AS ratios were deemed to be bound by AtGRP7 *in vivo*, plus the altered AS ratio of these transcripts did not occur in AtGRP7-RQ-ox mutant transgenic lines indicating that AtGRP7 affects the AS of these transcripts by directly interacting with them (Streitner *et al*., 2012). AtGRP7 is rhythmically activated by the endogenous circadian clock (Staiger & Apel, 1999), with global profiling of transgenic plants that constitutively overexpressed *AtGRP7* (AtGRP7-ox) showing enrichment of genes controlled by the biological clock (Streitner, Hennig, Korneli, & Staiger, 2010). Transcripts responsive to abiotic and biotic stress were enriched in AtGRP7-ox transgenic lines, with transcripts encoding two of the pathogenesis-related proteins (PR1 and PR2) being elevated in AtGRP7-ox plants but not in AtGRP7-RQ-ox transgenic plants, indicating that at least for some of the enriched transcripts direct binding of AtGRP7 is involved in their regulation (Streitner *et al*., 2010). Further evidence of the involvement of AtGRP7 in plant immunity was demonstrated by Fu *et al* (2007) who found that Arabidopsis KO mutants of *AtGRP7* were more

susceptible to *Pseudomonas syringae* pv. tomato (*Pst*) than wild-type plants. These experiments taken together indicated that some hnRNPs are involved in the plant defence response by regulating AS of defence-related genes, potentially via direct binding of the pre-mRNA.

1.4 The plant defence response

Due to their sessile nature plants are continually surrounded by a multitude of pathogenic organisms, however relatively few organisms can effectively infect or colonise any one particular plant species. Plants have evolved a complex multi-layered defence mechanism. The first lines of defence are preformed physical and chemical barriers that help prevent potentially pathogenic organisms from gaining entry to the plant. If an organism can circumvent these barriers, plants rely entirely on the innate immune response because they lack specialised immune cells or organs, transferring perception of a pathogen into an adaptive response.

Plants have two main defence mechanisms against potential pathogens that work alongside each other depending upon the type of microbial attack. Firstly the basal plant defence responses, such as patterned triggered immunity (PTI), effectively fend off the majority of potential plant pathogens via detection of microbe-associated molecular patterns (MAMPs) or damage-associated molecular patterns (DAMPs) via pattern recognition receptors (PRRs). In response, pathogens have evolved to produce virulence proteins (effectors) that suppress the basal plant defences, which can suppress plant defence response and actively manipulate the host environment for the pathogen's benefit. As part of an evolutionary arms race, plants developed a gene-for-gene resistance response, whereby intracellular plant disease resistance (R) proteins can recognise specific pathogen effector molecules activating effector-triggered immunity (ETI); in response, the effector recognized by an R protein can be lost or modified enabling the potential pathogen to colonise the plant (Figure 8).



Bent AF, Mackey D. 2007.
Annu. Rev. Phytopathol. 45:399–436

Figure 8. Model for the evolution of bacterial resistance in plants. (a) PTI is triggered by PRR such as receptor-like kinases (RLKs) recognising PAMPS or DAMPS produced by potential pathogens. PTI included MAP kinase cascades resulting in transcriptional reprogramming and effective plant defence. b) Pathogenic bacteria can suppress the plants basal immune responses using effector proteins, in this example delivered via type III secretion system (TTSS), resulting in an accumulation of the potential pathogen in the plant apoplast. (c) In a gene for gene response plants can evolve R genes that produce plant resistance proteins such as a TIR-NB-LRR protein. These R proteins detect effector activity triggering ETI restoring the plants resistance. (d) In an evolutionary arms race potential pathogens either lose effectors and replace them with an additional effector protein or modify existing effectors so they are no longer detected by plant R proteins, resulting in the plant being susceptible to the pathogen. Figure reproduced from Bent *et al* 2007.

Plant pathogens that are able to colonise plants have acquired effectors to collectively overcome PTI and ETI in their host plants (J. Zhang *et al*., 2010b).

This thesis predominantly focuses on how AS impacts on plant defence of *B. cinerea*; no known ETI response has been demonstrated for *B. cinerea*, however details of pathogen detection, mechanisms of PTI and effectors produced by *B. cinerea* have been investigated.

1.4.1 *B. cinerea* and the plant defence response

Plants detect infection of *B. cinerea* utilising receptor-like kinases (RLKs) and receptor-like proteins (RLP) that then associate with other RLKs or cytoplasmic kinases to transduce the signal to the mitogen activated protein kinases (MAPKs), triggering a kinase cascade. This in turn can activate transcription factors resulting in activation of defence responsive genes (Figure 9).

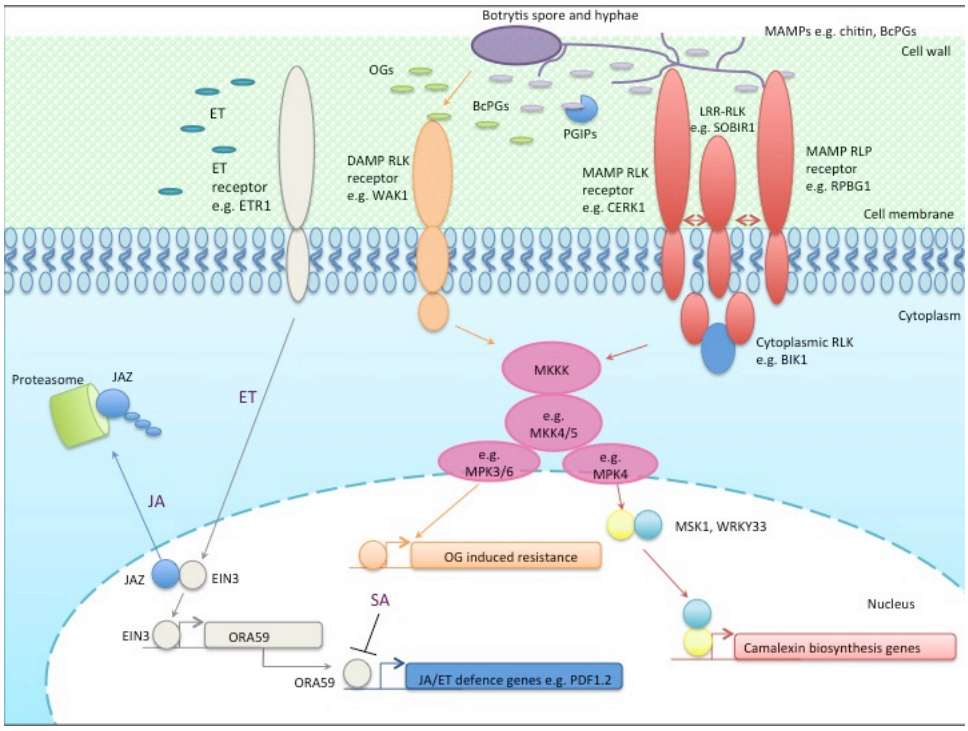


Figure 9. Overview of plant response to *B. cinerea* infection. RLKs and RLPs perceive infection and associate with other RLKS and/or cytoplasmic kinases, transducing the signal to MAPKKK. This triggers a MAP kinase cascade causing the activation of transcription factors, resulting in changes of gene expression. Crosstalk between phyotohormones, ethylene (*ET*), salicylic acid (*SA*) and jasmonic acid (*JA*) helps to regulate the defence response. → indicates positive regulation and ┤ indicates negative regulation. Diagram reproduced from Windram *et al* 2015.

Feeding into this defence response are phytohormones that have been shown to be involved in regulating the plant defence response via signalling crosstalk.

1.4.1.1 Detection of B. cinerea

Multiple MAMPS and DAMPs have been shown to be involved in the detection of *B. cinerea* by Arabidopsi. For example Polygalacturonase (PG), which has been shown to act as both a PAMP and a DAMP. It acts as a PAMP if it is the presence of the protein independent of its enzymatic activity that triggers PTI. PG can also degrade pectin in host cell walls, resulting in the production of oligogalacturonides (OGs) that can act as DAMPs and trigger PTI (Ferrari *et al*., 2007; Poinssot *et al*., 2003). The most well studied PAMP in necrotrophic fungal pathogens is chitin, a component of fungal cell walls that is not found in plants; it is a classical fungal MAMP that triggers plant defence responses, with plant plasma membrane binding sites for chitin being identified in a variety of plant species (Okada *et al*., 2002).

The first chitin receptor was identified by Kaku *et al* (2006) in rice (*Oryza sativa*), this receptor (Chitin Elicitor Binding Protein [CEBiP]) is a LysM domain-containing RLP, it lacks an intracellular kinase domain indicating that it may work in conjunction with another membrane-bound receptor/s. Evidence that this receptor plays a critical role in the perception and transduction of chitin comes from *CEBiP*-specific knockdown by RNA interference cell lines (*CEBiP-RNAi*). These cell lines display a suppression of ROS generated by chitin. In addition, microarray analysis of these cell lines showed that the transcriptional reprogramming elicited by chitin was affected with a large number of genes that were usually suppressed or induced in response to chitin either not responding, or the response being much weaker compared to non-transformed cell lines (Kaku *et al*., 2006). Further evidence that there is an additional chitin receptor, other than CEBiP, comes from these microarray experiments, which although they displayed a suppression of transcriptional reprogramming, the expression pattern of some elicitor-responsive genes were not affected in the CEBiP-RNAi cell lines (Kaku *et al*., 2006). Based on this evidence that indicated there was a "partner protein" to CEBiP required for membrane signalling in response to chitin, Miya *et al* (2007) searched for additional RLKs in Arabidopsis and identified a LysM-containing RLK, CERK1 (chitin elicitor receptor kinase) that is essential for chitin

elicitor signalling. CERK1 contains three LysM motifs in the extracellular domain with an intracellular Ser/Thr kinase domain with autophosphorylation/myelin basic protein kinase activity, implying that CERK1 is a PRR that detects the fungal MAMP chitin in plants (Miya *et al*., 2007). CERK1, also identified as LysM-RLK1 (LYK1) by Wan *et al* (2008), has been shown to be important in plant defence response to fungal pathogens. Miya *et al* (2007) showed that mutants of this gene (*cerk1-1 (Ds-transposon) and cerk1-2 (T-DNA)*) completely lost the ability to respond to chitin and initiate defence responses. The effects included impairment of chitin elicitor-induced activation of MAPK (demonstrated in *cerk1-1)*, ROS generation (demonstrated in *cerk1-1 and cerk1-2)* and transcriptional reprogramming (demonstrated in *cerk1-1*). In addition Wan *et al* (2008) showed that in T-DNA mutants induction of chitin responsive genes were blocked. Mutant lines of *cerk1* are shown to have increased susceptibility to certain fungal pathogens, but not to bacterial ones. Miya *et al* (2007) demonstrated that *cerk1-2* weakly effected the resistance of Arabidopsis to the necrotrophic fungus *Alternaria brassicicola* (an incompatible fungal pathogen) but not with the compatible fungus *Colletotrichum higginsianum*. In addition Wan *et al* (2008) found a moderate increase in susceptibility to the fungal pathogens *A. brassicicola* and *Erysiphe cichoracearum* (a biotrophic powdery mildew fungal pathogen). These experiments demonstrate that CERK1 is essential for chitin signalling and the induction of plant innate immunity in Arabidopsis. In rice cells, a LysM RLK with 54% sequence identity in amino acid sequences to Arabidopsis CERK1, OsCERK1, is required for chitin signalling where it is hypothesized to form the sandwich-like receptor complex CEBiP-OsCERK1 in response to chitin (Hayafune *et al*., 2014; Shimizu *et al*., 2010).

In Arabidopsis, upon interaction with Chitin, homodimerization of AtCERK1 occurs and the intracellular kinase domain is activated (T. Liu *et al*., 2012; Petutschnig *et al*., 2010). There is evidence that in Arabidopsis, AtCERK1 may also interact with other proteins to detect chitin; CERK1 has a low binding affinity for chitin, and mutations in CERK1 that are predicted to block binding did not block chitin-induced CERK1

phosphorylation (Petutschnig *et al*., 2010). Arabidopsis contains three closely related CEBiP homologues, LYM1-3, however Wan *et al* (2012) showed that mutants of these genes (either singly or in a triple mutant) did not alter the expression of chitin responsive genes indicating that, unlike in rice, CEBiP-like proteins do not play a role in chitin perception in Arabidopsis. However Narusaka *et al* (2013) demonstrated that *lym2* loss-of-function transgenic plants showed an increase in susceptibility to *A. brassicicola*, implying that there is a LYM2 dependent but CERK1 independent disease resistance pathway against fungal pathogens in Arabidopsis.

In addition to CERK1, there are four other lysine motif receptor-like kinase proteins in Arabidopsis; Wan *et al* (2012) and Cao *et al* (2014) showed using loss-of-function mutants of LYKs that LYK4 and LYK5 are important for chitin signalling and plant innate immunity in Arabidopsis. Cao *et al* (2014) proposed that AtLYK5 is the primary chitin receptor in Arabidopsis and that it forms a chitin-inducible complex resulting in AtCERK1 kinase activation. Evidence for this is that AtLYK5 binds chitin with a higher affinity than CERK1. However an *in vitro* kinase assay showed that the kinase domain of LYK5 is inactive (Cao *et al*., 2014). Although the kinase domain of LYK5 is inactive experiments Utilising a transgenic loss-of-function LYK5 line (*Atlyk5-2*) that expressed either wild-type LYK5, LYK5 that lacked a kinase domain, or LYK5 that had a mutated kinase domain predicted to disrupt ATP binding ability, under the native promoter showed that the mutant lacking the kinase domain could not complement the *Atlyk5-2* mutant phenotype, whereas the other two lines could, implying that although the kinase domain is not required for chitin-induced signalling it does have a function (Cao *et al*., 2014). The AtLYK5 mutant that lacked the kinase domain did not co-immunoprecipitate with CERK1 (the other two transgenic lines did) indicating that the kinase domain of AtLYK5 is necessary for association with CERK1 (Cao *et al*., 2014). Cao *et al* (2014) propose that chitin signalling is transduced downstream through AtCERK1 kinase activity requiring LYK5 binding; LYK5 exists as a homodimer in cells in the absence of chitin. Binding of chitin to LYK5 leads to its association with CERK1, dimerization and kinase activation.

A variety of proteins have been shown to modulate chitin signalling via direct interaction with CERK1, for example, LysM RLK1-interacting kinase 1 (LIK1) was shown to be a CERK1-Interacting Kinase that can regulate plant immune responses in Arabidopsis (Le *et al*., 2014). Le *et al* (2014) propose that LIK1 has a negative role in chitin mediated response because *lik1* mutant plants produce more ROS in response to chitin treatment and they have increased phosphorylation of MPK3 and MPK6. In addition the mutant plants have an increased susceptibility to the necrotrophic fungal pathogen *S. sclerotiorum* and a have a decrease in expression of key genes in the Jasmonate (JA) /ethylene (ET) signalling pathway implying that they positively regulate the necrotrophic defence pathway (Le *et al*., 2014). LIK1 immunoprecipitates with CERK1 before chitin treatment. However this interaction decreases 30 minutes after treatment. This led Le *et al* (2014) to propose a model whereby LIK1 interaction with CERK1 keeps the receptor in an inactive state in the absence of chitin to prevent unrequired activation of innate immunity which would be detrimental to plant growth. In response to chitin, dissociation of LIK1 from CERK1 would occur shortly after chitin binding enabling the inhibition to be relieved and chitin-mediated signalling to be transduced downstream through AtCERK1 kinase activity (Le *et al*., 2014).

1.4.1.2 Transmitting the signals downstream

Downstream signalling modules are required to transduce the extracellular stimuli detected by PRR into intracellular responses such as the generation of ROS, callose deposition and transcriptional reprogramming.

1.4.1.2.1 The receptor-like cytoplasmic kinases

One of the RLK superfamily lacks an extracellular domain, the receptor-like cytoplasmic kinases (RLCKs); these RLCKs are directly activated by PRRs such as CERK1 and have been shown to play an important role in PTI signal transduction in both Arabidopsis and rice (Shinya *et al*., 2014; K. Yamaguchi *et al*., 2013). In rice, the presence of chitin causes the RLCK OsRLCK185 to be directly phosphorylated by OsCERK1 and its subsequent dissociation from the OsCERK1 complex. RNAi silenced

transgenic suspension-cultured cell lines of OsRLCK185 display a suppression of expression of defence genes, ROS production and activation of MAP kinases indicating that OsRLCK185 is involved in signal transduction from OsCERK1 to the MAP kinase cascade (K. Yamaguchi *et al*., 2013). An Arabidopsis ortholog of OsRLCK185, PBL27, has been shown to be an immediate downstream component of CERK1 that plays a role in the regulation of chitin-induced immunity (Shinya *et al*., 2014). The fusion protein, PBL27–GFP, was transiently expressed in *N. benthamiana* where it localised to the plasma membrane. This combined with co-immunoprecipitation and bimolecular fluorescence complementation assays using transient expression in *N. benthamiana*, showed an interaction between PBL27 and CERK1 indicating that PBL27 interacts with CERK1 at the plasma membrane (Shinya *et al*., 2014). Evidence that this interaction with CERK1 results in phosphorylation of PBL27 was determined using differently tagged CERK1 kinase domains. PBL27 that were recombinantly expressed in *E. coli* and tested for the ability of CERK1 to phosphorylate PBL27. Further evidence of this is that in *cerk1* loss-of-function mutants chitin-induced phosphorylation of PBL27 is lost (Shinya *et al*., 2014). In *pbl27* loss-of-function mutants, there is a reduction in chitin-induced MAP Kinase activation, callose deposition and expression of two defence-related genes. However chitin-induced generation of ROS was not reduced, and flg22-induced MAP Kinase activation, callose deposition, defence gene expression and flg22-induced generation of ROS was not reduced in *pbl27* mutants indicating that PBL27 selectively regulates chitin-induced defence responses (Shinya *et al*., 2014). To investigate the molecular basis of the selectivity between CERK1 and FLS2/BAK1 *in vitro* phosphorylation experiments were carried out which showed that PBL27 is phosphorylated by CERK1 but only very weakly by BAK1, compared to BAK1's phosphorylation of BIK1 (a known downstream target of BAK1), indicating that there are different uses of RLCKs by different PRRs (Shinya *et al*., 2014). These authors showed that CERK1 also phosphorylate BIK1, but much more weakly that it did PBL27, indicating that BIK1 does play a role in chitin signalling in Arabidopsis, potentially by integrating the signalling from multiple PRR as proposed by Zhang *et al*

2010 (Shinya *et al*., 2014; J. Zhang, Li, *et al*., 2010a). Shinya *et al* 2014 propose that PBL27 and BIK1 regulate overlapping but partially different MAMP induced signalling pathways in Arabidopsis with both RLCKs being involved in MAMP-induced callose deposit, but only BIK1 being involved in ROS generation and only PBL27 being involved in the MAPK cascade.

Combining information from the above experiments by Shinya *et al* (2014) with that from Cao *et al* (2014) and Le *et al* (2014) we can propose a possible model of how CERK1 chitin-mediated signalling occurs (Figure 10).

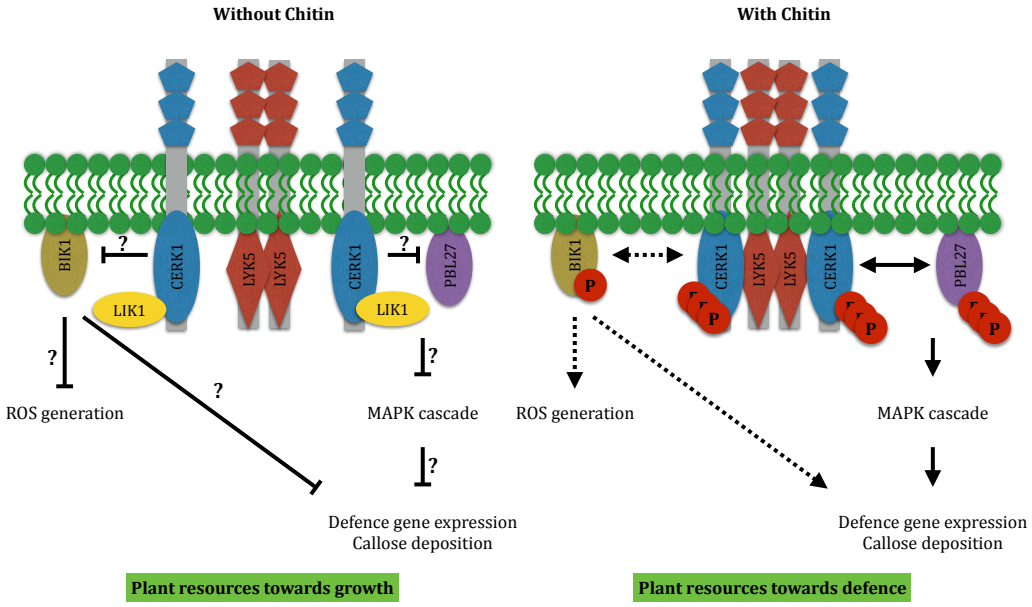


Figure 10. A possible model for Chitin perception and signalling in Arabidopsis. In the absence of Chitin (left hand panel), LYK5 form homodimers and LIK1 binds to CERK1 and inhibits (through an unknown mechanism) ROS generation, the MAP kinase cascade and JA/ET defence-related genes, potentially targeting the plant resources towards growth. In response to chitin (Right hand panel) LIK1 dissociates from CERK1, CERK1 and LYK5 bind to form a possible tetramer to mediate chitin signalling, with AtCERK1 being phosphorylated at the same time. In turn, AtCERK1 phosphorylates PBL27 and weakly phosphorylates BIK1. Phosphorylated PBL27 activates the MAP kinase cascade resulting in change of expression of defence-related genes and callose deposition, with BIK1 phosphorylation resulting in the generation of ROS and callose deposition, but not activation of the MAP kinase cascade, potentially resulting in plant resources being directed towards plant defence. In this model, AtLYK5 serves as a chitin perception protein, but AtCERK1 is responsible for chitin signalling transduction due to the inactive kinase domain of AtLYK5. → indicates positive regulation and ⊣ indicates negative regulation. Figure compiled from information and figures by Cao *et al* (2014), Shinya *et al* (2014) and Le *et al* (2014).

In the absence of chitin, LYK5 forms homodimers, and LIK1 binds to CERK1 to keep the receptor in the inactive state, resulting in plant resources being directed towards growth (left hand panel of Figure 10). In the presence of chitin LIK1 dissociates from CERK1 enabling the receptor to become active, CERK1 forms a possible tetramer with LKY5 and concurrently is phosphorylated. In turn CERK1 phosphorylates PBL27, which activates the MAP kinase cascade and callose deposition, plus weakly phosphorylating BIK1, which activates ROS generation and callose deposition (right hand panel of Figure 10). Interestingly PBL27 regulates chitin-induced activation of MAPKs but not the accumulation of ROS, which is different from BIK1 (Shinya *et al*., 2014).

In addition to the PRR-mediated perception of chitin, *B. cinerea* can also be perceived by plants detecting PGs with , for example, the PRR AtRLP42 recognizes fungal PGs (L. Zhang *et al*., 2014a). AtRPL42 is a RLP and as such needs to work in conjunction with one or more RLKs to transduce ligand binding into intracellular signalling (Zipfel, 2014). It is hypothesis that AtRLP42 utilizes BAK1 and/or SOBIR1 as co-receptors to transduce the signal through to the map kinase cascade (Mbengue *et al*., 2016).

1.4.1.2.2 The MAP kinase cascade

MAP kinase cascades are utilised by all eukaryotes to transfer signals from external and internal stimuli, such as perception of pathogens by PAMPs and DAMPs, minimally consisting of a MAP3K that phosphorylates a MAP2K that then phosphorylates a MAPK. They have been shown to be important for a variety of plant functions including playing an important role in the plant defence response, with the three most well studied MAPKs, the MPK3, MPK4 and MPK6 cascades being activated in response to pathogens (Colcombet & Hirt, 2008).

The first complete pathogen responsive plant MAP kinase cascade was identified by Asai *et al* (2002) utilising an Arabidopsis protoplast assay based on the induction of transcription of the early-defence genes WRKY29 and FRK1 (FLG22-INDUCED

RECEPTOR-LIKE KINASE 1) by the bacterial PAMP flg22. The MAP kinase cascade they identified acted downstream of the FLS2/BAK1 receptor complex and contained one MAP3K (MEKK1), two MAP2Ks (MKK4 and MKK5) and two MAP3Ks (MPK3 and MPK6) that activated the transcription factors WRKY22 and WRKY29 (Asai *et al*., 2002). Agrobacterium-mediated transiently transformed Arabidopsis leaves expressing constitutively active MEKK1, MKK4, MKK5 or wild-type WRKY29 displayed enhanced resistance to *Pst* and *B. cinerea* indicating that signalling events initiated by a broad range of pathogens may converge into a conserved MAP kinase cascade (Asai *et al*., 2002). The constitutively active MKK4, MKK5 mutants activate MPK3 and MPK6 (Asai *et al*., 2002). This information, combined with evidence from Miya *et al* (2007) that demonstrated that chitin activates MPK3 and MPK6 and that this activation is completely lost in the *cerk1* mutant and with work by Shinya *et al* (2014) that showed PBL27 loss-of-function mutant Arabidopsis plants had a suppression of the activation of MPK3 and MPK6, indicates that this signalling cascade may be downstream of a variety of PRR complexes that are activated by a range of PAMPs. These experiments also indicate that the MPK3/6 cascade is a positive regulator of the defence response. Initially it was thought that the MPK4 cascade was a negative regulator of the defence response because loss-of-function mutants constitutively express defence responses including enhanced salicylic acid (SA), increased resistance to pathogens (*Pst* DC3000 and *Peronospora parasitica*) and expression of defence-related genes (Petersen *et al*., 2000). However, Kong *et al* (2012) showed that the MPK4 cascade negatively regulated MEKK2 which is a positive regulator of SUMM2 (an R protein); Zhang *et al* (2012) proposed that the inhibition of SUMM2 by MPK4 via MEKK2 could mask a positive role for MPK4 in basal defence responses. Utilising loss-of-function mutants including *summ2* they determined that MPK4 positively regulates basal defences whilst negatively regulating immunity specified by the R gene SUMM2. Disruption of components of the MPK4 cascade would thus activate SUMM2 resulting in activation of downstream responses explaining the constitutive expression of defence responses in single *mpk4* loss-of-function mutants (Z. Zhang *et al*., 2012).

The output of the MAP kinase signalling cascade activated in response to *B. cinerea* is phosphorylation of a wide range of MAPK substrates, including enzymes, and transcription factors that can activate defence-related genes including those that are involved in the synthesis of anti-microbial compounds (such as phytoalexins and glucosinolates) and those involved in the synthesis of phytohormones (Han *et al*., 2010; G. Mao *et al*., 2011; J. Xu *et al*., 2016).

The regulation of the phytoalexin camalexin, the accumulation of which has been shown to be a key factor in influencing susceptibility to *B. cinerea*, is regulated by the activation of MAP kinase cascades in response to pathogens (Denby *et al*., 2004; Qiu *et al*., 2008). Qui *et al* (2008) showed using immunoprecipitation assays and examination of the phosphorylation status using SDS-PAGE, that in uninfected leaves, MPK4, MKS1 and WRKY33 are bound in a complex. In response to infection with the pathogen *Pst*, or treatment with flg22, MPK4 is activated and WRKY33 is released. Utilising ChIP they showed that WRKY33 is recruited to the promoter of *pad3* (a gene known to encode an enzyme that catalyses the final step in camalexin biosynthesis (Schuhegger *et al*., 2006)) *in vivo* in response to both pathogen infection and flg22 treatment. The MPK3/MPK6 cascade has also been shown to play a key role in the synthesis and regulation of camalexin via transcriptional regulation of the biosynthetic genes in response to infection by *B. cinerea* (Ren *et al*., 2008). Mao *et al* (2011) determined that the transcription factor WRKY33 was also involved in the reprogramming of the camalexin biosynthetic genes via the MPK3/MPK6 signalling pathway. Evidence for this is that in the *wrky33* loss-of-function mutant Arabidopsis plants both gain-of-function MPK3/MPK6 (where MPK3/MPK6 is always signalling) and *B. cinerea* camalexin induction was compromised. In addition they showed that WRKY33 is a substrate of MPK3/MPK6 using an *in vitro* MAPK phosphorylation assay with an *in vivo* phosphoprotein mobility shift assay used to demonstrate that WRKY33 is phosphorylated by MPK3 and MPK6 in response to *B. cinerea* infection (G. Mao *et al*., 2011). Taken together these experiments indicate that the MPK3/MPK6 signalling cascade leads to direct phosphorylation of WRKY33 which in

turn causes transcriptional reprogramming of camalexin biosynthetic genes effecting the camalexin-mediated defence response.

Using similar techniques this group also detected that the transcription factor ETHYLENE RESPONSE FACTOR6 (ERF6) is also a substrate of MPK3/MPK6, and that phosphorylation of ERF6 by MPK3 and MPK6 occurs *in vivo* in response to *B. cinerea* infection. In addition, utilising a phospho-mimicking ERF6, they demonstrated that some defence-related genes were constitutively activate, including the plant defensin genes *PDF1.1* and *PDF1.2* (genes encoding peptides with antimicrobial properties), resulting in enhanced resistance to *B. cinerea* (X. Meng *et al*., 2013). The glucosinolate, indole glucosinolate (IGS), has also been shown to be regulated by the MPK3/MPK6 signalling cascade, via the transcription factor ERF6. Xu *et al* (2016) demonstrated using gain and loss-of-function transgenic plants that the *B. cinerea*– induced MPK3/MPK6 cascade was involved in regulation of expression of two transcription factors involved in the IGS biosynthetic pathway, MYB51 and MYB122, as well as regulating the expression of CYP81F2 and IGMT1/IGMT2 (enzymes involved in the IGS biosynthetic pathway) via ERF6 (J. Xu *et al*., 2016). In addition ChIP-RQ-RT PCR experiments showed that CYP81F2 and IGMT1/IGMT2 are direct targets of ERF6 (J. Xu *et al*., 2016). Taken together, these experiments demonstrate that transcriptional reprogramming in response to *B. cinerea* induced MPK3/MPK6 activation involves a network of transcription factors, some of which directly interact with defence genes and others that activate other transcription factors, to regulate the synthesis of anti-microbial compounds in response to *B. cinerea* infection. This type of transcriptional reprogramming via transcription factors is established as an essential control mechanism in plant defence (Birkenbihl & Somssich, 2011).

1.4.1.3 Transcriptional reprogramming and the plant defence response

Defence against *B cinerea* results in large-scale transcriptional programming, as shown by whole-genome transcriptional profiling of Arabidopsis leaves infected with *B. cinerea* utilising microarrays (Abuqamar *et al*., 2006; Windram *et al*., 2012). These

experiments revealed that there are complex regulatory networks involved in plant defence against *B. cinerea*.

1.4.1.3.1 Gene regulatory networks mediate the plant defence response

Work by Windram *et al* (2012) investigating the transcriptional reprogramming of Arabidopsis in response to *B. cinerea* shows that multiple transcription factors are involved in the plant defence response, and their putative regulation may be inferred by network modelling. They utilised a high-resolution time series that captured gene expression profiles spanning the 48 hours post infection. The underlying transcriptional networks were modelled via inferring regulatory relationships between genes based upon correlation of clustered gene expression profiles. The resulting network contained some nodes consisting of genes with transcription factor binding motifs in their promoter regions, and in several cases the upstream cluster contained a transcription factor that could bind the motif; this indicates that these inferred relationships may enable the prediction of regulatory interactions (Windram *et al*., 2012). The network consisted of multiple different transcription factors interacting with a vast array of different genes, highlighting the complex regulatory networks involved in plant defence (Windram *et al*., 2012). Massive transcriptional reprogramming was detected in this time series, with one third of Arabidopsis genes undergoing transcriptional reprogramming in 48 hours post infection, covering a vast range of diverse functions from down-regulation of genes associated with photosynthesis, potentially for nitrogen to be re-allocated to the synthesis of new defence proteins, to the up-regulation of genes associated with autophagy (Windram *et al*., 2012). Genes associated with defence phytohormones (ET, JA, SA and ABA) also showed massive transcriptional reprogramming and, interestingly, a distinct chronological order of gene ontological enrichment associated with the different phytohormones was seen. Ethylene biosynthesis occurred early in infection (14 hpi), shortly followed by responses to ethylene and JA at 16 hpi, enrichment of terms associated with ABA catabolism and negative regulation of ABA signalling occurred later (approximately 20 and 22 hpi

respectively), with auxin biosynthesis also occurring late at 22 hpi. (Windram *et al*., 2012).

The Arabidopsis genome contains over 2000 genes that encode transcription factors, which can be divided into over 50 families with approximately half being found only in plant species (Shinozaki & Sakakibara, 2009. In section "1.4.1.2.1" members of three of some of the most influential transcription families in regards to plant defence were briefly discussed, the WRKY, the MYB and the ERF transcription families. The WRKY transcription factors have been implicated as one of the most important regulators of transcriptional reprogramming associated with plants response to pathogens, with the majority of the 74 Arabidopsis *WRKY* genes being activated by defence-related stimuli (Eulgem, 2006).

1.4.1.3.2 The WRKY transcription factors and *B. cinerea* infection

The WRKY transcription factor family is characterised by a highly conserved WRKYGQK motif that can interact with the W-box (a DNA cis-regulatory element sequence, (T)TGAC[C/T]) affecting the transcription of genes (Rushton *et al*., 1996). They have been shown to be important in plant pathogen responses with many being implicated in the defence response of Arabidopsis to *B. cinerea* including WRKY3, 4, 8, 18, 28, 33, 40, 60 and 70 (Abuqamar *et al*., 2006; L. Chen *et al*., 2010; Lai *et al*., 2008; Windram *et al*., 2012; L.-T. Wu *et al*., 2011; X. Xu *et al*., 2006; Zheng *et al*., 2006).

In addition to its role in camalexin biosynthesis as discussed in section "1.4.1.2.2", WRKY33 has also been shown to be involved in regulating autophagy, callose deposition and phytohormone signalling (Birkenbihl *et al*., 2012; Lai *et al*., 2011; S. Liu *et al*., 2015b; T. H. Reyes *et al*., 2015).

WRKY transcription factors do not work in isolation to bring about plant defence via the modulation of phytohormone pathways. Shim *et al* (2013) showed that *AtMYB44* can modulate two antagonistic phytohormone pathways by activating SA-mediated

defences and repressing JA-mediated defences via directly binding to the promoter region of *WRKY70*; they propose that *AtMYB44* modulates and fine-tunes this network of genes via the activation of WRKY70, causing activation of SA-mediated genes (including PR1, which is activated by NPR1) and suppression of JA-responsive gene expression (including PDF1.2, which can be activated by COI1) (Figure 11).

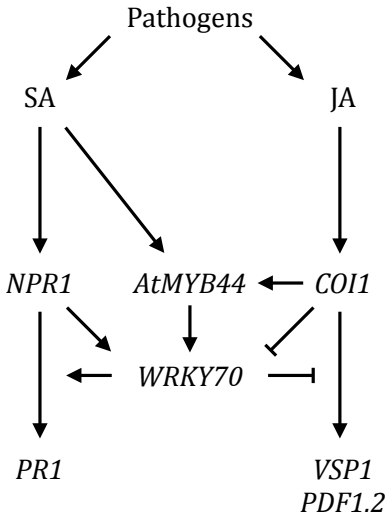


Figure 11. The role of AtMYB44 in the antagonistic interaction between SA- JA-mediated defence signalling. Detection of pathogens can activate the SA- or JA- mediated pathways; in the SA-mediated pathway AtMYB44 is activated in an NPR1 independent manner, and in JA-mediated pathway it is activated in a COI1 dependent manner. Activated *AtMYB44* regulates the activation of the SA marker gene *PR1* and the suppression of the JA mediated genes *VSP1* and *PDF1.2* via directly binding to the promotor of *WRYK70*. JA activated AtMYB44 acts as a negative regulator and fine-tunes the JA signalling. COI1 also directly interacts with WRKY70. Diagram reproduced from Shim *et al* (2013).

1.4.1.3.3 The MYB transcription factors and *B. cinerea* infection

The most prominent members of MYB transcription factors in respect of plant defence to *B. cinerea* are MYB51 and AtMYB44, discussed above, and BOS1 (AtMYB108), a well-known transcription factor that is involved in the response to biotic and abiotic stresses in Arabidopsis (Mengiste *et al*., 2003; Vailleau *et al*., 2002). Transgenic loss-of-function *bos1* Arabidopsis mutants show increased susceptibility to *B. cinerea* as well as an impaired ability to cope with the abiotic stresses water deficit, increased salinity, and oxidative stress (Mengiste *et al*., 2003). Genetic analysis revealed that *BOS1* expression increases in response to *B. cinerea* in wild-

type plants and that this response was reduced and delayed in *coi1* loss-of-function mutants, which are insensitive to JA, but remains unaltered in *ein2* and *NahG* loss-of-function mutants, which are insensitive to ethylene and unable to accumulate SA respectively, indicating that the increased in susceptibility of *bos1* mutants to *B. cinerea* is due to BOS1 interacting with the jasmonate signalling pathway (Mengiste *et al*., 2003).

1.4.1.3.4 The AP2/ERF transcription factors and *B. cinerea* infection

The AP2/ERF transcription factor family is one of the largest in Arabidopsis with over 140 members, with the ERF subfamily particularly being involved in regulating biotic stress (Tsuda & Somssich, 2015). ORA59 has been shown to be directly bound by WRY33 (Birkenbihl *et al*., 2012), and, using genetic experiments utilising loss-of-function and gain-of-function transgenic lines to be an essential integrator of the JA and ET signalling pathways where it acts as a positive regulator by binding GCC-box motifs in the promoters of target genes (Pré *et al*., 2008; Zarei *et al*., 2011). This highlights the fact that the transcription factor families work together to bring about transcriptional reprogramming often by mediating crosstalk between phytohormone signalling transduction pathways. ERF6 has also been shown to play a positive (but redundant with ERF5) role in the *B. cinerea* defence likely due to its role in mediating the antagonistic crosstalk between the JA/ET and SA pathways (Moffat *et al*., 2012). In response to *B. cinerea,* ERF6 can be phosphorylated by the MPK3/MPK6 kinase cascade and in turn regulates two MYB transcription factors indicating that transcription factors form complex networks with each other, and these networks are often involved in mediating the crosstalk between hormone signalling pathways.

1.4.1.4 Phytohormones and plant defence response to B. cinerea

Phytohormones are small signalling molecules that occur in low concentrations that are vital for the regulation of plant growth, development, and reproduction as well as playing an important role in the plant defence response (Pieterse *et al*., 2009). They regulate the plant defence response against both biotic and abiotic stresses via synergistic and antagonistic actions, referred to as signalling crosstalk (Fujita *et al*.,

2006). Traditionally the key defence related phytohormones were SA, JA and ET, with JA and ET viewed as the hormones associated with defence against necrotrophic pathogens, such as *B. cinerea*, and SA being involved in defence against biotrophic organisms (Glazebrook, 2005). However, experiments utilising the quadruple mutant *dde2*/*ein2*/*pad4*/*sid2* in Arabidopsis, showed that immunity to the necrotrophic pathogen *A. brassicicola* was severely compromised compared to single mutants indicating that SA, PAD4, JA and ET, all positively contribute to the defence response of necrotrophs, with network modelling showing a complex interwoven network of the core phytohormones involved in the response to pathogens (Tsuda *et al*., 2009).

JA has long been known to positively regulate plant defence against *B. cinerea* with many genetic studies enabling a relatively clear picture of how the JA-mediated signalling pathway brings about resistance. One of the first and possibly the most important JA-mediated signalling mutant found to be more susceptible to *B. cinerea* was *coi1-1*, a loss-of-function mutant that was show to be insensitive to JA due to a defect in the CORONATINE INSENSITIVE1 (COI1) protein (Thomma *et al*., 1998; Xie, Feys, James, Nieto-Rostro, & Turner, 1998). COI1 is a F-box protein that has been shown to form part of the Skp1/Cullin/F-box complex (SCF^{COI1}) *in vivo* (Devoto *et al*., 2002; L. Xu, 2002). The SCF^{COI1} complex is a E3 ubiquitin ligase that targets proteins, specified by the F-box, for degradation by the proteasome. Using a combination of molecular, genetic and biochemical techniques it has been shown JA-Ile is the active hormone and in response to its presence the JAZ proteins are degraded by the proteasome. In addition, the JAZ proteins also interact with the bHLM transcription factor MYC2 (a positive regulator of JA-mediated signalling) (Chini *et al*., 2007; Thines *et al*., 2007; Yan *et al*., 2007).

The JAZ family consists of 12 original members plus JAZ13 (which was added in 2015). They are characterised by three conserved domains, the N terminal domain, which contains an EAR motif in six of the JAZ proteins (JAZ1, JAZ2, JAZ5, JAZ6, JAZ7

and JAZ8), a ZINC-FINGER EXPRESSED IN INFLORESCENCE MERISTEM (ZIM) domain containing a TIFY (JAZ1-JAZ12) or NAFY (JAZ13) motif, and a C-terminal Jas domain (Thireault *et al*., 2015; Wager & Browse, 2012). In order to identify the core JA-mediated signalling and to find novel JAZ interactions, Pauwels *et al* (2010) utilised a tandem affinity purification (TAP) assay (both in the presence and absence of activate JA) that resulted in isolating NINJA (Novel Interactor of JAZ). This was confirmed, and the domains of interactions were determined, using Y2H and bimolecular fluorescence complementation. Following this they utilized the same techniques to investigate what interacted with NINJA, identifying that the protein TOPLESS (TPL) interacts with NINJA (Pauwels *et al*., 2010). Results from these key experiments and previous literature led Pauwels *et al* (2010) to propose the following model: in the absence of JA-Ile, MYC transcription factors interact with the Jas domain of the JAZ proteins, the JAZ proteins interact through their TIFY motif with domain C of NINJA, with the EAR motif of NINJA interacting with the TPL co-repressors resulting in the repression of transcription of early JA responsive genes. In the presence of JA-Ile, JAZ proteins interact with the ubiquitin ligase SCF^{COI1} leading to ubiquitination and proteosomal JAZ degradation via the 26S proteasome pathway; this is followed by the release of the NINJA–TPL complex from the MYC transcription factors enabling the activation of jasmonate-responsive gene expression (Figure 12) (Pauwels *et al*., 2010).

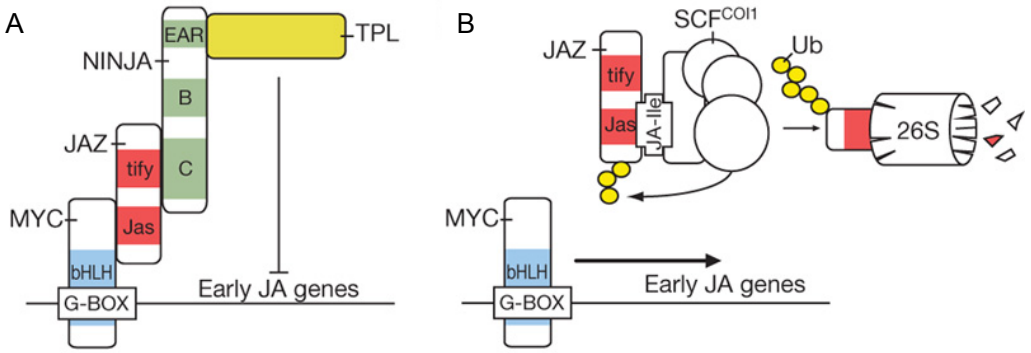


Figure 12. Model for a general function of TPL proteins in plant hormone signalling. A. In the absence of the active form of jasmonate (JA), JA-Ile, MYC factors interact with the JAS domain of the JAZ proteins, which in turn interact through their TIFY domain with the C domain of NINJA, the NINJA EAR motif interacts with TOPLESS (TPL) resulting in the repression of transcription of early JA genes. B In the presence of JA–Ile, the hormone facilitates interaction between JAZ proteins and the ubiquitin ligase SCF^{COI1} resulting in JAZ proteins being ubiquitinated (yellow circles) and subsequently degraded in the 26S proteasome, releasing transcription factors from inhibition and activating JA-responsive gene transcription. Adapted from (Pauwels *et al*., 2010).

The JAZ proteins have also been linked to the crosstalk between the JA and ET-mediated pathways, where they physically interact with two transcription factors ETHYLENE INSENSITIVE3 (EIN3) and EIN3-LIKE1 (EIL1) and recruit I6, resulting in the transcriptional activity of EIN3 and EIL1 being repressed (Zhu *et al*., 2011). In the ET-mediated signalling pathway, perception of ET represses a Raf-like kinase CONSTITUTIVE ETHYLENE RESPONSE 1 (CTR1), which negatively regulates downstream ET-mediated signalling events (Kieber, Rothenberg, Roman, Feldmann, & Ecker, 1993). Downstream of CTR1 is the essential positive regulator of ET-mediated signalling ETHYLENE INSENSITIVE 2 (EIN2), with EIN3 and EIL1 being downstream of this (Alonso *et al*., 1999; Chao *et al*., 1997). Zhu *et al* (2011) propose that EIN3/EIL1 acts as a key integration node where the presence of ET is required for the stabilisation of EIN3/EIL1 and JA is needed for EIN3/EIL1 to be released from the JAZ repressors to enable effective transcriptional reprogramming for defence against necrotrophic pathogens, potentially explaining the synergistic and mutually dependant roles of ET and JA in plant defence (Zhu *et al*., 2011). In contrast, the crosstalk between the JA and SA-mediated pathways is thought to occur downstream of the SCF^{COI1}-JAZ complex, via regulating the accumulation of

transcription factors, such as ORA59, that target GCC-box motifs in JA-responsive genes (Van der Does *et al*., 2013).

The experiments above highlight how complex and interwoven the phytohormone signalling pathways are, consisting of a large degree of crosstalk between pathways, which can be either synergistic or antagonistic, much of which mediates and can be mediated by transcription factors helping to fine-tune the defence responsive transcriptional reprogramming. In addition to undergoing massive transcriptional changes in response to pathogens, post-transcriptional changes can radically affect the plants ability to mount an effective plant defence response; here we will particularly focus on differential alternative splicing.

1.4.2 Differential Alternative Splicing and the plant defence response

Differential alternative splicing (DAS) is when splicing patterns alter depending upon cell-type, developmental-stage or environmental cues. DAS of R genes was one of the first examples of defence related DAS in plants. Several R genes are regulated by AS of their transcripts, this may produce proteins with different combinations of functional domains or truncated proteins that contain only a subset of the domains found in full-length transcripts (Marone *et al*., 2013). In a number of examples, both versions of the transcripts have been shown to be important for effective pathogen defence (Dangl & Jones, 2001; Dinesh-Kumar & Baker, 2000; Jens Staal, 2008; Michael Weaver *et al*., 2006; X.-C. Zhang & Gassmann, 2003). Perhaps the best known example is the N resistance gene that confers resistance to tobacco mosaic virus (TMV) in tobacco. Dinesh-Kumar *et al* (2000) have shown that the N resistance gene encodes two isoforms, N_S that can encode the full N protein and N_L which encodes a truncated protein N^{tr} (Dinesh-Kumar & Baker, 2000). Transgenic plants that were only capable of expressing the N protein but not the N^{tr} protein did not exhibit full resistance to TMV, whereas transgenic plants containing constructs capable of producing both the N_S and N_L transcripts did (Dinesh-Kumar & Baker, 2000). Interestingly constructs that were predicted to produce N_S and N_L in a 1:1 ratio also did not exhibit full resistance to TMV leading to the supposition that the

ratio or timing of the two transcripts and their protein isoforms may play a crucial role in the defence response, evidence for this hypothesis comes from measuring the ratios of the transcripts over the course of infection (Dinesh-Kumar & Baker, 2000). Prior to infection, the N_s transcript dominated with the ratio of N_S :N_L being approximately 28:1. At 6 hpi the alternative transcript that encodes the truncated transcript (N_L) was the most dominant with a 1:23 ratio, with initial level ratios returning by 9 hpi (Dinesh-Kumar & Baker, 2000).

In Arabidopsis, the disease resistant protein gene RPS4, also produces multiple transcripts, some of which have truncated open reading frames via AS of introns flanking exon 3 and a cryptic intron within exon 3. Zhang and Gassmann 2003 showed using stably expressed intron-deprived transgene in susceptible Arabidopsis plants that RPS4-mediated resistance requires both full length and transcripts with truncated open reading frames. To determine if the ratio of truncated to full-length transcript was required to confirm resistance (as is the case for N gene mediated resistance to TMV) or if dynamic alternative splicing is required, they quantified the five experimentally verified alternatively spliced RPS4 transcripts and concluded that only the one with intron 3 retained is transiently up-regulated in response to *Pst* DC3000 indicating that RPS4-mediated resistance is under dynamic regulation (X.-C. Zhang & Gassmann, 2007).

It is unclear how truncated proteins play a positive role in plant defence. However there are a few theories with the most popular being that they either act as adapter molecules for downstream signalling, or alleviate self-inhibition of the full length proteins (Gassmann, 2008; Marone *et al*., 2013). Meyers *et al* (2002) proposed that IR-NBS (TN) proteins, proteins that have a similar structure to R proteins but lack the LRR domain, could be plant analogues of small TIR-adapter proteins such as MyD88 and Mal, that function in mammalian innate immune responses. Zhang and Gassmann propose that truncated proteins transcribed from AS genes with the TN structure may also function this way (Zhang & Gassmann, 2003; 2007). The theory

proposed by Zhang and Gassmann (2003), whereby truncated proteins may work to alleviate self-inhibition, could potentially work by full-length R proteins forming dimers, with the C-terminal negatively regulate function. The truncated proteins lack the C-terminal, thus when binding with full length proteins they would alleviate this self-inhibition, and as part of a population containing resistance complexes, facilitate activation (X.-C. Zhang & Gassmann, 2003). Consistent with this theory is that truncated constructs of RPS4 by themselves could not bring about resistance and, of particular note, is the fact that both *RPS4* and *N* have a C-terminal non LRR domain, while L6 (a gene that confers resistance to flax rust in flax) does not have a C-terminal non LRR domain and no evidence that alternative transcripts for L6 were required for resistance was found (X.-C. Zhang & Gassmann, 2003).

1.4.2.1 PRR and alternative splicing

The PRR CERK1 detects the classical fungal MAMP chitin in plants (Miya *et al*., 2007), and the splicing of CERK1 has been shown to play an important role in plant immunity (Z. Zhang *et al*., 2014b). Zhang *et al* 2014 demonstrated SUA and RSN2 are required for the proper splicing of CERK1 and PAMP triggered immunity, loss-of-function mutants of these genes show an increase in the retention of intron one of *CERK1*, accompanied by a decrease in chitin-mediated ROS induction (flg22-mediated induction was comparable to wild-type), and an increase in susceptibility to a non-pathogenic *Pst* (DC3000hrcC) (Z. Zhang *et al*., 2014b). This evidence that *SUA* and *RSN2* can regulate the alternative splicing of *CERK1*, which in turn can affect immunity, raises the question of whether *CERK1* is DAS in response to pathogen infection and whether could this be a potential regulatory mechanism of plant defence.

Other potential PRRs also undergo AS in plants. For example members of the LRR-RLKs, many of which are known to act as PRR (Mastrangelo, Marone, Laidò, De Leonardis, & De Vita, 2012; Shiu & Bleecker, 2001), are known to be regulated in part by AS with a large proportion of LRR-RLKs being shown to undergo AS in Arabidopsis (Gou *et al*., 2010; Mastrangelo *et al*., 2012). Determining the extent to

which this family undergoes differential alternative splicing (DAS) in response to *B. cinerea* infection could help to elucidate how DAS regulates the plant defence response.

1.4.2.2 The MOS4 associated complex and plant defence

The MAC is involved in pathogen defence and required for the proper splicing of R genes, potentially representing a link between the defence response and DAS.

The core components of the MAC are involved in the plant defence response with loss-of-function mutants for *AtCDC5*, *MOS4*, *PRL1* and the double loss-of-function mutant *mac3a mac3b* displaying increased susceptibility *to Pseudomonas syringae* pv *maculicola*, *Pst,* and *Hyaloperonospora arabidopsidis* Noco2, indicating that they are essential for plant innate immunity (Monaghan *et al*., 2009; Palma *et al*., 2007). I have previously shown that the double mutant *mac3a mac3b* also has increased susceptibility to the necrotrophic fungal pathogen *B. cinerea* (Foster, 2012). In fact, the three core members, MOS4, AtCDC5, and PRL1 were identified by utilising *snc1* as an autoimmune model; *snc1* mutants display enhanced resistance to a variety of pathogens which reverts to wild-type resistance in double *snc1 mos4* mutants (Palma *et al*., 2007).

In addition to this, the core components of the MAC are essential for the proper splicing of defence-related R genes. Experiments utilising loss-of-function mutants in transgenic Arabidopsis plants have shown that *mos4*, *Atcdc5* and the functionally redundant *mac3a* and *mac3b* are required for the 49orrectect splicing of *SNC1* and *RPS4* (Monaghan *et al*., 2009; F. Xu *et al*., 2012). The correct splicing of *RPS4* and *SNC1* is also affected by another modifier of *snc1*, MOS12, which is thought to be a component of the MAC because it co-immunoprecipitates with MOS4 (F. Xu *et al*., 2012). However, MOS14, another modifier of *snc1* which has not been shown to co-immunopreciptitate with MOS4 and thus is not thought to be a component of the MAC, also affects the proper splicing of SNC1 and RPS4 (S. Xu *et al*., 2011), indicating that the splicing of R genes is effected by many proteins and not just those that are

part of the MAC. Additional components of the MAC are predicted to effect splicing of R genes due to their homology to PRP19 or CDC5L components. For example MAC5a and MAC5b are predicted to effect splicing of R genes because their human counterpart RBM22 interacts with the U6 snRNP of the spliceosome (S. Xu *et al*., 2011). Although it is known that some components of the MAC affect splicing of the R genes, SNC1 and RPS4, how they affect other R genes and components of the plant defence response is unknown. Genome-wide splicing investigations into the effects of *mos4-1* transgenic plants will provide information about the global effects of this component of the splicing machinery, helping to determine what range of genes splicing patterns are affected and if this is specific to defence-related genes.

1.4.2.3 DAS of SR proteins in plant defence

MOS14 is involved in the transportation of SR proteins into the nucleus and as mentioned above is required for the proper splicing of the R genes, *SNC1* and *RPS4* (S. Xu *et al*., 2011). MOS14 has been shown function as a Transportin-SR (TRN_SR); this is a nuclear import receptor for SR proteins. The transgenic loss-of-function mutant Arabidopsis line, *mos14-1*, displays an impairment of localisation of several SR proteins into the nucleus and has compromised resistance mediated by *SNC1* and *RPS4* in addition to altering splicing patterns of *SNC1* and *RPS4* (S. Xu *et al*., 2011). This indicates that import of SR proteins by MOS14 is required for the proper splicing of *SNC1* and *RPS4* and that the effective splicing of these genes plays an important function in plant defence.

Multiple genes encoding SR proteins have been shown to undergo DAS in response to abiotic stresses, and *SR30* undergoes DAS in response to *Pst* infection (Ding *et al*., 2014; Filichkin *et al*., 2010; Gulledge *et al*., 2012; Howard *et al*., 2013; Lazar & Goodman, 2000; Palusa *et al*., 2007; Tanabe *et al*., 2007). However to our knowledge no SR proteins have been detected that undergo DAS in plants in response to necrotrophic pathogens.

1.4.2.4 hnRNP, splicing and the defence response

In addition to SR proteins displaying plant defence mediated DAS other *trans*-acting factors such as hnRNPs have also been implicated in the plant defence response. As mentioned in section "1.2.2.2.2" AtGRP7 is involved in the plant pathogen defence response, as shown by experiments utilising loss-of-function mutants being more susceptible to *Pst* and gain-of-function mutants having increased levels of the defence-related genes, *PR1* and *PR2* (Z. Q. Fu *et al*., 2007; Streitner *et al*., 2010), potentially by regulating AS of pre-mRNA of target genes. They have been shown to be alternatively spliced, however it has not been shown if they undergo DAS or differential gene expression (DGE) in response to pathogens. Another hnRNP protein that has been linked to plant defence is LIF2. In *lif2* mutants basal levels of both the defence marker genes *PR1* and *PDF1.2* (which are associated with SA and JA signalling pathways respectively) are altered. In addition genes relating to biosynthesis and signalling of the SA and JA pathways are affected. However there is no change in the hormone levels (Le Roux *et al*., 2014). Reduced susceptibility to both *Pst* and *B. cinerea* was observed in the loss-of-function transgenic lines, and the *lif2-1 sid2-2* double mutant was also less susceptible to *Pst* indicating that the reduced susceptibility is SA independent (*salicylic acid induction deficient2* (*sid2*) encodes a gene involved in SA biosynthesis) (Le Roux *et al*., 2014). It has been proposed that the *lif2* loss-of-function mutant displays a basal primed defence state, indicating that LIF2 may be involved in regulating the conflicting defence and development pathways (Le Roux *et al*., 2014). Investigating if these hnRNPs undergo DAS or are differentially expressed in response to pathogen attack may help to further expand our knowledge on how the hnRNPs help regulate plant defence.

1.4.2.5 AS-coupled NMD and plant defence

Stress can have a dramatic effect on the alternative splicing of pre-mRNAs. This coupled to NMD may have a profound effect on the plant defence mechanisms. A variety of experiments have demonstrated that NMD plays a crucial role in the plant defence response, potentially indicating that AS-coupled NMD may be a crucial component of plant responses to pathogens.

NMD-defective mutants have been shown to have enhanced resistance to *Pst*, increased levels of SA and elevated pathogenesis-related transcripts, indicating that the NMD pathway is involved in the plant defence response to biotrophic pathogens (H.-J. Jeong *et al*., 2011; Rayson *et al*., 2012). Experiments carried out by Riehs-Kearnan *et al* (2012) in Arabidopsis indicate that impaired NMD leads to activated ETI. They showed that *smg7-1* loss-of-function mutants display characteristics of lesion mimic mutants, indicating that they have an enhanced hypersensitive response. This is usually found in response to the activation of R-gene/s whose signals are transduced through either the NDR1 protein or the PAD4/ EDS1 complex. Inactivation of PAD4 in *smg7-1* suppresses these characteristics, resulting in SA and pathogenesis-related transcripts being restored to normal. However *smg7-1 pad4* mutant plants still display elevated levels of the defensin PDF1.2, which is regulated by the JA signalling pathway, indicating that there is still an on-going defence signal (Riehs-Kearnan *et al*., 2012). Thus, impairment of NMD in Arabidopsis results in aberrant pathogen response signalling, implying that NMD plays a role in regulating plant immunity. Gloggnitzer *et al* (2014) showed that NMD operates post-transcriptionally to modulate defence activation thresholds of a subset of R proteins (the TNLs, which are NB_LRR proteins with a Toll/human Interleukin-1 receptor (TIR) domain in the N terminal sequence) and that transgenic *smg7* loss-of-function mutant Arabidopsis plants infected with *Pst* displayed enhanced resistance. However, the PTI responses were unaltered, implying that NMD-regulation of the defence response acts downstream of early PTI mechanisms and is likely to contribute to ETI mechanisms through modulation of TNLs (Gloggnitzer *et al*., 2014). They also demonstrated that the isoform of the R gene RPS6 generated from an excision of a cryptic intron is up-regulated in NMD deficient mutants. This combined with work by Kim *et al* (2009) that showed that *RPS6* is AS upon infection with *Pst* implies that DAS-coupled NMD may play a regulatory role in plant defence of biotrophic organisms (Gloggnitzer *et al*., 2014; S. H. Kim *et al*., 2009). In addition Shi *et al* (2012) determined that the NMD genes *UPF1*, *UPF2* and *UPF3* were required for a variety of stress responses including a biotic stress, *Pst* infection, and the abiotic

stress, wounding; they found that these genes where involved in modulating JA levels and transcript abundances of two JA signalling marker genes *LOX2* and *VSP2* (C. Shi *et al*., 2012). This combined with the work of Riehs-Kearnan *et al* (2012) that shows *smg7-1 pad4* transgenic plants have elevated levels of the defensin PDF1.2, indicate that NMD regulation is also involved in plant defence of *B. cinerea*. One possibility is that receptors that trigger PTI in response to *B. cinerea* may be regulated by DAS coupled to NMD, potentially modulating defence activation thresholds in a similar manner to those shown by Gloggnitzer *et al* 2014 in R gene activation of ETI (Gloggnitzer *et al*., 2014). However further work would be required to demonstrate this.

Taken together these experiments indicated that NMD is involved in regulating the plant defence response. Investigations into genome-wide DAS of transcripts in response to *B. cinerea* and predicting which transcripts will be subject to NMD will further expand our knowledge of how DAS coupled to NMD helps to fine-tune the plant defence response.

Most work on AS and the defence response has been carried out using *Pst*, a hemi biotroph. Thus investigating a necrotrophic pathogen such as *B. cinerea* will give us information regarding whether AS plays a role in the defence response, and if so, what aspects it is involved in. This will help further our understanding of the plant defence response to pathogens.

Plant defence against pathogens has proven to involve a complex network of genes that can be altered at both the transcriptional and post-transcriptional level. Although plant defence has been implied to involve DAS, very few global studies have been undertaken to investigate how global-wide AS patterns alter in response to pathogen infection. Such investigations will enable a greater understanding of plant defence, which may in the future potentially be able to be used to breed plants that are more resistant to disease, thus improving food security.

Aims and objectives

The overarching aim of this thesis is to investigate the importance of DAS in plant defence and to begin to elucidate how it is regulated, with particular focus on the defence response to the necrotrophic fungal pathogen *B. cinerea*. In order to achieve this the following research objectives will be met:

1. Demonstrate if differential splicing is important in the plant defence response by
    a. Identifying a number of splicing events that are regulated in an infection-specific manner
    b. Determine if a subset of genes with infection-mediated DAS have a defence phenotype
    c. Investigate LRR-RLK to ascertain if DAS could potentially be involved in the regulation of PTI
2. Determine a robust set of Arabidopsis genes that are differentially spliced in response to *B. cinerea* by
    a. Critically evaluating RNA-Seq analysis techniques and developing a pipeline/s for use in analysing RNA-Seq data
    b. Identifying Arabidopsis genes that undergo *B. cinerea* mediated DAS in multiple biological samples
3. Begin to determine if defence-mediated DAS is pathogen-specific
4. Identify potential spliceosomal components that are involved in plant defence and could regulate *B. cinerea* mediated DAS by determining if components of the MAC

a. Are linked to *B. cinerea* defence, i.e. do MAC loss-of-function mutants have altered defence phenotype compared to wild-type
b. Affect *B. cinerea* mediated DAS of gene/s
c. Affect DAS at the genome-wide level

5. Begin to elucidate the "splicing code" by identifying DNA motifs associated with *B. cinerea* mediated DAS

2 Chapter 2: Material and Methods

2.1 Experimental methods

2.1.1 Plant lines

The Arabidopsis mutants and wild-type plants (Col0) used in this study are in the Columbia genetic background. All T-DNA knockout lines were obtained from the Nottingham Arabidopsis Stock Centre (NASC) unless otherwise stated in (Table 3)**Error! Reference source not found.**.

Table 3. Arabidopsis lines. Column one identifies the type of experiments the lines were used in, column two and three give the gene identifier and name respectively. Column four and five give the line used and where it was obtained. Red highlighted cells indicate predicted homozygote lethal and yellow highlighted cells indicate potentially more than one T-DNA insertion.

Type	AGI	Name	Line	Obtained from
Hpa border		ws-eds1	ws-eds-1	(Parker *et al*., 1996)
Pst and *Hpa* positive control		NahG	NahG-1	(Delaney *et al*., 1994)
B. cinerea positive control	AT3G06490	bos1	Bos-1	(Mengiste, Chen, Salmeron, & Dietrich, 2003)
MAC	Double mutant	MAC3A MAC3B	mac3a mac3b	(Monaghan *et al*., 2009)
MAC	AT1G04510	MAC3A	SALK_143098C	NASC
MAC	AT1G04511	MAC3A	SALK_089300C	NASC
MAC	AT1G06220		SALK_025707	NASC
MAC	AT1G07360	MAC5A	Mac5a-1	(Monaghan *et al*., 2009)
MAC	AT1G07360	MAC5A	SALK_142085C	NASC
MAC	AT1G09770	AtCDC5/MAC1	Atcdc5-2	(Monaghan *et al*., 2009)
MAC	AT1G10580		SALK_068763C	NASC
MAC	AT1G15200		SALK_000102C	NASC
MAC	AT1G20960	EMB1507	SAIL_408_D03	NASC
MAC	AT1G34290	EMB2733/ESP3	SALK_040466C	NASC
MAC	AT1G77180	SKIP	SAIL_681_H11	NASC
MAC	AT1G80070	SUS2/EMB177	SAIL_905_A07	NASC
MAC	AT2G33340	MAC3B	SALK_144856C	NASC
MAC	AT2G33340	MAC3B	SALK_066437C	NASC
MAC	AT2G38770	EMB2765	SALK_141518C	NASC
MAC	AT2G38770	EMB2765	SALK_041627C	NASC

MAC	AT2G43770		SALK_049286	NASC
MAC	AT3G15730	PLDALPHA1	SALK_053786	NASC
MAC	AT3G18165	MOS4	Mos4-1	(Monaghan *et al*., 2009)
MAC	AT3G18790		SALK_102990C	NASC
MAC	AT3G20820		SALK_119747C	NASC
MAC	AT3G20820		SAIL_154_A05	NASC
MAC	AT3G60190	ADL1E	SALK_044933C	NASC
MAC	AT4G15900	PRL1/MAC2	RIF_18	(Monaghan *et al*., 2009)
MAC	AT4G19410		SALK_093502C	NASC
MAC	AT5G28740		SALK_006170C	NASC
MAC	AT5G41770		SALK_030126C	NASC
MAC	AT5G42080	ADL1	SALK_069077	NASC
MAC	AT5G64270		SALK_135626C	NASC
LRR-RLK DAS	AT4G39270		SALK_072419C	NASC
LRR-RLK DAS	AT4G39270		SALK_051652C	NASC
LRR-RLK DAS	AT4G39270		SALK_047268C	NASC

2.1.1.1 Confirmation of T-DNA insertion

For all lines obtained from NASC, the presence of the T-DNA insertion was confirmed by DNA extraction on randomly selected 10-day-old seedlings, using the Red Extract-N-Amp kit as per manufacturers instructions. A standard PCR was run and agarose gel electrophoresis was used to visualise the DNA on a 1.5% agarose gel run at 80V. For SALK and SAIL lines the left border primer 5'-GCGTGGACCGCTTGCTGCAACT-3' and 5'-GCCTTTTCAGAAATGGATAAATAGCCTTGCTTCC-3' respectively were used in conjunction with the right and left primer specific to the gene to determine homozygosity as detailed in Table 4. Five repeats per line were carried out; for those lines that had no homozygous batches individual plants were tested at three weeks (ten plants per line).

Table 4. Primers sequences for homozygosity test.

AGI	Line	Right Primer (5'-3')	Left Primer (5'-3')
AT1G045	SALK_14309	ATGAGAATACTACTTTTCG	CAAACAATGCTGTGTGACA
AT1G045	SALK_08930	ATGTGGTAATCAGGATGTC	TCCATTGTGTCCACCAAAC
AT1G062	SALK_02570	CTTGTCCTTGAGGACAGTA	GGCTAACCTAAGTAGAGGT
AT1G073	SALK_14208	GGGATTTCGCCATTCTAAG	CCGCAGCTGTAATCAAATC
AT1G105	SALK_06876	CGGATTGGCATTGATCCAT	CCGAAAGACCATCGTTCTT

AT1G152	SALK_00010	GGAGACAAGAAGTGAGTG	CAGTGAAGAGACATCTGTC
AT1G209	SAIL_408_D	CGTTGGGCTTAGGAAGTAT	CTTATCCAGAAGCAGTTGG
AT1G342	SALK_04046	CCGAATCAAACAACACTCG	CAGGTTGACTCGATGTTTG
AT1G771	SAIL_681_H	GGGTAGGAGTGAAAGTGT	GGATTCGTGAACGGAAGA
AT1G800	SAIL_905_A	CAGCTCTTGTGCTGATGTTC	CAGCCCAAACACTTGTATG
AT2G333	SALK_14485	TGTTGTGATCGAGGATATG	TGCAATGTTTTCGGCTTTAT
AT2G333	SALK_06643	GAATGGAACCTTATCAAGA	CACGAATCTGATATTCGTG
AT2G387	SALK_14151	CACAGCTTGTCTTATGGAC	CTAGTTGGGCGAAAAGCTT
AT2G387	SALK_04162	GGATGTTGTCTGCTCAAAG	CTTTGGAAGAGCAAGGCA
AT2G437	SALK_04928	GAAACCCAATTTGTCAGCC	GTTTCCCAGTCTCAACATCC
AT3G157	SALK_05378	CCATTGGAGCTACCCTTATC	GAAGTCATCATGGTGGACT
AT3G187	SALK_10299	CTACCTCCATCTTCACTTGG	CAGGATAACGAGAAGGAG
AT3G208	SALK_11974	CGAGGCCAAAGTGTCTAAT	GATCAAACGGTGAACCTAC
AT3G208	SAIL_154_A	CGAGGCCAAAGTGTCTAAT	GATCAAACGGTGAACCTAC
AT3G601	SALK_04493	CGTATCTAACCGCTGAGTT	CGCACAAGTTTGACAAGTC
AT4G194	SALK_09350	GTGGCAAGCAAAGCACTAA	CGCCTAAGAGTATGTACAC
AT5G287	SALK_00617	GGATATGGATCTCGACGTT	CTCTGCAGATTCTTGCCAA
AT5G417	SALK_03012	CTGAAGGAAGAACGAGCT	GCCCTACCATCTAGTCTTAC
AT5G420	SALK_06907	GTGTTCCCATTCACCTTAGC	CCGCAATCATGTCGACATT
AT5G642	SALK_13562	CGGACTTTCGGCTATATGT	GTGTCCAGATTCGCATCAT
AT4G392	SALK_07241	GAGGAAGACACAACGGTG	AGCTCCAGAAGTATCTTGC
AT4G392	SALK_05165	GAGGAAGACACAACGGTG	AGCTCCAGAAGTATCTTGC
AT4G392	SALK_04726	CTGCAGTTGGAGGGTCTAT	ACAGTAGCAAGCTTGTTTG

2.1.1.2 Confirmation of knockdown

To confirm that lines are significantly knocked down / knocked out, RNA was extracted from leaf 8 of four-week-old soil grown plants; infected leaves were used for *AT4G39270* because this increases its expression. The RNA was converted into cDNA using SUPERSCRIPT™ II RNase H Reverse Transcriptase (Invitrogen) as per manufacturer's instructions. The cDNA was used as templates in a real-time quantitative reverse transcription PCR (RQ-RT PCR). Samples were diluted 1 in 10 with DEPC-treated water before being run in triplicate on a Roche LightCycler 480 using PrecisionPlus™ by PrimerDesign as per manufactures instructions. Gene expression was compared using the primers detailed in Table 5, and normalized to *AtPUX1* (AT3G27310), a gene with transcripts levels that have previously been shown not to change in response to *B. cinerea* infection (Windram *et al*., 2012) using primers 5'-TTTTTACCGCCTTTTGGCTA-3' and 3'-ATGTTGCCTCCAATGTGTGA-5'.

Table 5. Primers sequences for confirmation of knock-out/ knock down of gene expression.

AGI	Line	Forward Primer (5'-3')	Reverse Primer (5'-3')
AT1G073	SALK_14208	CTTACATGATGCCTCCACAC	CATTCTCTGTGGGTCCATT
AT1G097	Atcdc5-2	CATAGCCATAGCCATGGAA	CAAAGTCAAGCTACGTCTC
AT2G387	SALK_14151	GCAAGATCTCTTCAGCTTCC	GTTTAGGGGTGTCTGAGAA
AT2G387	SALK_04162	GCAAGATCTCTTCAGCTTCC	GTTTAGGGGTGTCTGAGAA
AT3G181	Mos4-1	GTATGAGCGTGTTAGAGCT	GGTTGTTCTGTCTCCAAAG
AT3G208	SALK_11974	CTGAAAGGACCAATACCGA	GATCAAACGGTGAACCTAC
AT3G208	SAIL_154_A	CTGAAAGGACCAATACCGA	GATCAAACGGTGAACCTAC
AT4G159	RIF_18	GGGACCTTGAGCAGAATAA	GTGTTGTCATGTCCTGAGA
AT4G392	SALK_07241	GAGGAAGACACAACGGTG	AGCTCCAGAAGTATCTTGC
AT4G392	SALK_05165	GAGGAAGACACAACGGTG	AGCTCCAGAAGTATCTTGC
AT4G392	SALK_04726	CTGCAGTTGGAGGGTCTAT	ACAGTAGCAAGCTTGTTTG

2.1.2 Plant growth

2.1.2.1 Soil grown plants

Arabidopsis seeds were stratified for 3 days in 0.1% agarose at 4°C in the dark before sowing onto Arabidopsis soil mix (Scotts Levingtons F2s compost:sand:fine grade vermiculite in a ratio of 6:1:1) containing intercept.

Plants were grown at 20°C, 350 ppm CO_2, 70% humidity, and light intensity of 100 μmol photons-m^{-2}·s^{-1}. Plants in long and short day conditions were under a 16:8-h and 10:14 light:dark cycle respectively. For all experiments except *Hpa* phenotyping, p24 trays were used to grow plants. If more than one line was required the lines were randomized between and within trays to prevent growth condition bias.

2.1.2.2 Agar grown plants

Seeds were sterilised and sown onto square agar trays containing Murashige and Skoog (MS) medium. Trays were placed upright in SANYO cabinet set on light setting 4, grown at 22°C, 10:14 light:dark cycle.

2.1.3 Pathogen growth of *B. cinerea*

B. cinerea strain pepper (Denby *et al*., 2004) was sub-cultured onto sterile tinned apricot halves in Petri dishes 2 weeks prior to use of the spores. Subcultures were incubated in the dark at 25°C. Spore inoculums were prepared by harvesting spores in water, filtration through microcloth to remove hyphae, and suspension in half-strength sterile grape juice to a concentration of 10^5 or 10^4 spores mL^{-1} as required.

2.1.4 Phenotype screens

2.1.4.1 B. cinerea

Twelve leaf 7 leaves of 28-day-old plants of the test lines plus one leaf seven of both Col0 and *bos1* were detached and placed on a bed of 0.8% agar in three propagator trays. A scale measure was placed between every other line. Ten leaves from each line was spotted with 5µl of 10^4 spores mL^{-1} spore suspension and the remaining two leaves with 5µl of 50% filtered grape juice 9 hours after subjective dawn. Trays were covered with lids, sealed and placed in a growth chamber with the same conditions as for plant growth. Photos of the duplicate screen trays were taken at 34, 48, 54 and 72 hours for *At4g39270* screen and 34, 48 and 72 hpi for the MAC screen using Nikon D50 Camera. The area of the lesion was determined using ImageJ.

2.1.4.2 B. cinerea growth

B. cinerea β-tubulin gene expression levels were used as a measure of the growth of *B. cinerea* in leaf 7 of four-week-old *Arabidopsis* plants. This was determined using RQ-RT PCR. RNA was extracted from individual leaves (1 leaf per biological replicate) using TRIzol (Invitrogen) and purified using the Qiagen RNeasy Mini Kit. The RNA quality and purity was evaluated using NanoDrop; the integrity of the RNA was determined using the Agilent 2100 BioAnalyser. All samples required a RNA Integrity Number (RIN) higher greater than five, this was used because that is the minimum RNA quality value recommendation by Schroeder *et al* (2006) for use with RQ-RT PCR. RNA exceeding this threshold then underwent DNase treatment using the Ambion Turbo DNase™ kit as per manufacturers instructions. 1 µg of RNA was converted into cDNA using SUPERSCRIPT™ II RNase H Reverse Transcriptase

(Invitrogen) as per manufacturer's instructions. The cDNA was used as templates in a RQ-RT PCR.

Samples were diluted 1 in 10 with DEPC-treated water before being run in triplicate on a Roche LightCycler 480 using PrecisionPlus™ by PrimerDesign as per manufactures instructions. β-tubulin gene expression levels were compared using the primers 5'-TTCCATGAAGGAGGTTGAGG-3' and 5'-TACCAACGAAGGTGGAGGAC-3', to *PUX1* expression using primers 5'-TTTTTACCGCCTTTTGGCTA-3' and 5'-ATGTTGCCTCCAATGTGTGA-3'.

2.1.4.3 Chitosan screen

Chitosan (obtained from SIGMA) was dissolved in 400 μl of 5% acetic acid and diluted to 150mg/l, 100mg/l and 50mg/l in 20 ml volumes with sterile H_2O. Mock inoculum consisted of 20ml H_2O and 400 μl of acetic acid. 28-day-old plants were spray-inoculated using an airbursh system attached to a compressor (GS, model AS18). 20ml of Chitosan solution was used per 24 plants (one tray) at a pressure of 1012 psi, the plants were allowed to dry for 30 minutes before being replaced in growth cabinet. Leaf 8 leaves were harvested 2hpi and snap frozen in liquid nitrogen.

2.1.4.4 Pseudomonas Syringe pv tomato DC3000

Pseudomonas syringae pv. Tomato (*Pst*) DC3000 was streaked onto Kings B agar plates supplemented with rifampicin (100 μg/ml) from frozen glycerol stocks, using a sterile loop and incubated at 28°C for 2 days. Single colonies were transferred to liquid Kings B medium supplemented with rifampicin (100 μg/ml) and incubated overnight in a shaking incubator (28°C, 220rpm). 20ml of inoculum was prepared per 24 plants at an OD600 = 0.1 and siluet L-77 (a silicone surfactant) was added to create a final concentration of 0.05%. Plants were spray-inoculated using an airbursh system attached to a compressor (GS, model AS18) at a pressure of 10-12 psi. $1cm^2$ leaf disc were obtained from separate plants (at 1, 2 and 3 dpi) and immediately placed into 1.5ml Eppendorf tubes containing 200 μl of $MgCl_2$. Sterile glass beads were added to the tubes and the samples were ground using a mixer mill (2X 30

seconds, 28 Hertz). 800 μl of $MgCl_2$ was added and the samples were serially diluted. 10μl of each dilution was plated onto Kings B agar and incubated (24 hours at 28°C) followed by a second incubation step (24 hours, room temperature). Colonies were manually counted and the number of colony forming units (CFU) calculated. *Nahg* ((Delaney *et al*., 1994)) was included in each screen as positive control.

2.1.4.5 Hyaloperonospora parasitica

Arabidopsis seeds were sown directly onto p40 trays containing Arabidopsis soil mix (Scotts Levingtons F2s compost:sand:fine grade vermiculite in a ratio of 6:1:1) containing intercept, and then stratified for 3 days in 0.1% agarose at 4°C in the dark. With each outside module of the tray containing ws-eds1 (a mutant with enhanced disease susceptibility) to create a border to aid with infection, test lines were randomly assigned an inside module. Plants were grown at 20°C, 350 ppm CO_2, 70% humidity, and light intensity of 100 μmol photons-m^{-2}·s^{-1} in 10:14 light:dark cycle.

Six modules of stock seedling inoculated with *Hyaloperonospora parasitica* (*Hpa*) *Nok1* were harvested into 50 ml falcon tubes and 20 ml of cold sterile H_2O added, this was filtered using microcloth to remove seedlings, and the concentrations adjusted to 30000 conidiospores per ml.

Two-week-old seedlings were spray-inoculated with a suspension of 30000 conidiospores per ml of *Hpa* isolate noks1 using an airbursh system attached to a compressor (GS, model AS18) at a pressure of 10-12 psi. Trays were covered with lids, sealed and placed in a growth chamber with the same conditions as for plant growth. At four days post infection the number of conidiophores per 15 seedlings was counted. *NahG* was included in each screen as a positive control.

2.1.4.6 Morphological

2.1.4.6.1 Soil based

Five-week-old plants were photographed using Nikon D50 Camera and the leaf area determined using ImageJ.

2.1.4.6.2 Agar based

Ten-day-old seedlings were photographed using Nikon D50 Camera and the root length determined using ImageJ. Seedling were then completely removed from agar and weighed to give dry weight.

2.1.5 Molecular biology techniques

2.1.5.1 Determining ratio of spliced intron relative to retained intron for At4g39270

For *B. cinerea* response, the seventh leaf of four-week-old plants were detached and placed on a bed of 0.8% agar in propagator trays. Leaves were inoculated with either 5µl droplets of *B. cinerea* inoculum at a concentration of 10^5 spores mL^{-1} or with mock inoculum. The trays were covered with lids, sealed and kept under the same conditions as for plant growth. For chitosan experiments see section "2.1.4.3" for sample preparation.

RNA was extracted and cDNA preparation was carried out as per section "2.1.4.2". The cDNA used as templates in a RQ-RT PCR were diluted 1 in 2 with DEPC treated water before being run in triplicate on a Roche LightCycler 480 using PrecisionPlus™ by PrimerDesign as per manufactures instructions. The following primers were utilised, forward for retained intron one 5'-CGTCGTCCACAGGTATGAAAAACC-3' and forward for spliced intron one 5'-CGTCCACAGGGATATTCAAGCAAG-3' the same reverse primer was used for both, 5'-GTCTGAGCAATCGGGCTATTTTCC-3'. The Pfaffl variation of the ΔΔcq method with a 1.74 correction was used to analyse the RQ-RT PCR data (Pfaffl, 2001).

2.1.5.2 RNA-Seq experimental procedures

Two leaves of number eight leaf from 28-day-old plants of each of the following lines, Col0, *Atcdc5*$_{KO}$ and *mos4*$_{KO}$ were detached and placed on a bed of 0.8% agar in three separate propagator trays. One Col0 leaf per tray was inoculated with five 5µl droplets of *B. cinerea* inoculum at a concentration of 10^5 spores mL^{-1} so that droplets were evenly spaced over the leaf. The remaining Col0 leaves were mock-inoculated

with five 5µl droplets of sterile half-strength grape juice. For the MAC mutants the amount of inoculum was administered in a size dependent manner worked out using the following equations:

$$\frac{Col0_{LeafArea}}{5} = x$$

$$\frac{KO_{LeafArea}}{x} = Number\ of\ Droplets$$

resulting in the *mos4*$_{KO}$ and *Atcdc5*$_{KO}$ leaves receiving three and two 5µl droplets respectively.

The trays were covered with lids, sealed and kept under the same conditions as for plant growth. Leaves were harvested at 24 hpi and snap frozen in liquid nitrogen. RNA was extracted using TRIzol (Invitrogen) and purified using the Qiagen RNeasy Mini Kit. The RNA quality and purity was evaluated using NanoDrop, and the integrity by the Agilent 2100 BioAnalyser, all samples required RIN of 7±0.4 to be carried forward for the RNA-Seq experiment and underwent DNase treatment using the Ambion Turbo DNase™ kit as per manufactures instructions. mRNA was isolated from the total RNA, fragmented and cDNA made as per the Illumina TruSeq® RNA Sample Preparation v2 Guide. Paired-end 100 bp sequencing was performed at the Wellcome Trust Centre for Human Genetics.

2.2 Analysis of RNA-Seq data

Quality assessment of the RNA-Seq raw reads was carried out with FastQC (version 0.10.1) on CyVerse (Goff *et al*., 2011). The Zhang *et al* (2015) annotation file, from now onwards referred to AtRTD, was utilised in the RNA-Seq analysis.

2.2.1 Alignments

2.2.1.1 Tophat2

RNA-Seq reads were aligned using Tophat2-PE in CyVerse (Tophat2 for paired end data) (Goff *et al*., 2011). The reference genome was TAIR10 and AtRTD was used for the reference annotations. Default settings were used with the exception of minimum and maximum intron size and Mate-Pair Inner distance. Tophat2 was

designed on mammalian systems, thus the minimum and maximum intron length is not appropriate for Arabidopsis. All intron lengths were obtained from TAIR10 and the range of intron size that accounted for 99.98% of the data was used for the minimum and maximum intron length, 15 to 5000 respectively. The mate-paired inner distance was calculated using the Picard value to be -35. Tophat2-PE utilised version 2.0.9 of Tophat and 2.1.0 of bowtie.

2.2.1.2 STAR

RNA-Seq reads were aligned to AtRTD transcriptome using STAR via CyVerse (Goff *et al*., 2011). The genome index was generated using STAR_2.4.0.1_GenerateGenomeIndex with default setting with the overhang set to 99. Then STAR_2.4.0.1 utilized the generated genome index to align reads, default settings were used with the exception of minimum and maximum intron length which was set the same as for Tophat2 (minimum 15, maximum 5000).

2.2.1.3 Subread

Subread 1.4.6 was run from the command line to align RNA-Seq reads. First the index was built, and then the reads were aligned. Example commands are below:

```
subread-buildindex -o my_AtRTD_index AtRTD.fasta
subread-align -i my_AtRTD_index -r Col0_M1_1 -R Col0_M1_2 -o
subread_AtRTD_results_M1.sam
```

2.2.2 Quantifiers

2.2.2.1 FeatureCount

FeatureCount 1.4.6 used the alignment outputs as input and was run from the command line with default settings. An example command would be:

```
FeatureCount –a AtRTD.gtf –t exon –g gene_id –o counts.txt
Col0_M1_subread.sam Col0_M2_subread.sam Col0_M3_subread.sam
Col0_I1_subread.sam Col0_I2_subread.sam Col0_I3_subread.sam
```

2.2.2.2 HTSeq

The output files from the alignment section were used as input for HTSeq-count 0.5.4 on CyVerse (Goff *et al*., 2011). The AtRTD gtf file was utilised in conjunction with default settings, the mode was set at intersect-nonempty, and the given feature was either gene or isoform depending upon whether detection of differentially expressed genes or differentially expressed isoforms were required.

2.2.2.3 Sailfish

Sailfish version 0.6.3 was used as an alignment-free quantification method. It was run from the command line with an index generated from the AtRTD reference transcript followed by quantification. An example command would be:

```
sailfish index -t AtRTD.fasta -o AtRTD_index -k 20
sailfish quant -i AtRTD_index -l "T=PE:O=><:S=U" -1 Col0_M1_1
-2 Col0_M1_2 -o Col0_M1
```

2.2.3 DEG and DEI algorithms

2.2.3.1 CuffDiff2

CuffDiff2 was utilised via CyVerse (Goff *et al*., 2011), with the reference annotation being AtRTD, reference genome TAIR10 and default settings; version 2.1.1 was used.

2.2.3.2 DESeq

DEseq was utilised via CyVerse (Goff *et al*., 2011), with default settings.

2.2.3.3 DESeq2

DESeq2 was run in R version 3.1.3. An example command can be found in section 7.1

2.2.3.4 edgeR

EdgeR was run in R version 3.1.3. An example command can be found in section 7.2

2.2.4 Event level algorithms

2.2.4.1 SUPPA pipelines utilising the Robinson and Smyth exact test

SUPPA version (1.1.0) was run in python 2.7 and used to generate the different AS events from AtRTD, see below for example command.

```
suppa.py generateEvents -i AtRTD.gtf -o
AtRTD_generatedEvents -e SE SS MX RI FL
```

Tab delimited text documents from the outputs in the section "2.2.2" were created using the transcript per million (TPM) scores and SUPPA was used to determine the percentage spliced in (PSI) of events. An example command where we determine the retained intron (RI) events from the Sailfish quantifier is given below:

```
suppa.py psiPerEvent – GeneratedEvents_RI.ioe –e
Sailfish_TPM.txt -o SUPPA_Sailfish_TPM_RI
```

This was repeated for all events and quantifiers; a generic usage of the program is given below:

```
suppa.py psiPerEvent --ioe-file <ioe-file> --expression-file
<expression-file> -o <output-file>
```

RI events where the average coverage was less than 10% of total transcript reads in the RI area in either the mock, infected or both samples were filtered out. To determine if the percentage spliced in (PSI) of events from the remaining data was statistically significantly different between mock and infected samples (i. e. did differential alternative splicing (DAS) occurr), the Robinson and Smyth exact test (Robinson & Smyth, 2008) was performed in R 3.1.3 with the Benjamini & Hochberg multiple testing correction method being applied.

2.2.4.2 ASTALAVISTA pipelines utilising Robinson and Smyth exact

ASTALAVISTA software was used to extract all events in the Zhang *et al* 2015 annotation file (R. Zhang *et al*., 2015)) for the four main AS events RI, A3SS, A5SS and SE and relative inclusion levels were determined.

To determine the relative inclusion levels, i.e. the PSI, of RIs both junction and body reads were utilised. To calculate the PSI the number of reads aligning within the intron (*Ireads*), were divided by the number of reads that spanned the splice junction (*SJreads*) plus those that aligned within the intron as per equation 1.

$$1) \qquad \frac{Ireads}{SJreads + Ireads} = PSI\_RI$$

RI events could potentially be due to partially processed transcripts that contain introns that are yet to be spliced out. To account for this, only those RI events where there was a minimum average coverage of 10% of total transcript reads in the RI area in either the mock, infected or both samples were analysed.

For the other events only the junction reads were utilised with the relative inclusion levels being calculated as the number of junction reads supporting the inclusion event, divided by the inclusion reads plus the number of reads supporting the exclusion event as per the equations in Alamancos *et al* 2015 (Alamancos, Pagès, Trincado, Bellora, & Eyras, 2015). Examples are given below:

2.2.4.2.1 Skipped Exon

For a given skipped exon event eg:
AT4G30850;SE:Chr4:15021211-15021703:15021781-15021879:-
The junction reads 15021211-15021879 would be supporting the exclusion event (this would correspond to the [thickStart] [thickEnd] being 15021211 and 15021703 in the Tophat2 junction.bed file respectively), and junction reads supporting the inclusion events would be 15021211-15021703 and 15021781-15021879. Thus in equation 2, reproduced from Alamancos *et al* (2015), C_{12} would be junction score of 15021211-15021703, C_{23} would be the junction score of 15021781-15021879, and C_{13} is the junction score of 15021211-15021879. Junction scores are obtained from either Tophat2 junction.bed file or STAR SJ.out.tab file, and the PSI worked out using equation 2, reproduced from Alamancos *et al* (2015).

$$2) \qquad PSI_{SE} = \frac{c_{12} + C_{23}}{2C_{13} + C_{12} + C_{23}}$$

2.2.4.2.2 Alternative 3’ splice site

For a given alternative 3’ splice site event, eg:
AT5G09450;A3:Chr5:2941898-2941988:2941898-2942018:+

The junction reads supporting the inclusion event (C_{12} in equation 3), are 2941898-2941988 with the junction reads supporting the exclusion events (C_{13} in equation 3) being 2941898-2942018. Junction scores are obtained from either Tophat2 junction.bed file or STAR SJ.out.tab file, and the PSI worked out using equation 3, reproduced from Alamancos *et al* (2015).

$$3) \qquad PSI_{A3SS} = \frac{C_{12}}{C_{12} + C_{13}}$$

2.2.4.2.3 Alternative 5′ splice site

For a given alternative 5′ splice site event, eg:

AT1G12550;A5:Chr1:4275095-4275307:4275054-4275307:+

The junction reads supporting the inclusion event (C_{21} in equation 4), are 4275095-4275307 with the junction reads supporting the exclusion events (C_{23} in equation 4) being 4275054-4275307. Junction scores are obtained from either the Tophat2 junction.bed file or STAR SJ.out.tab file, and the PSI worked out using equation 4, reproduced from Alamancos *et al* (2015).

$$4) \qquad PSI_{A5SS} = \frac{C_{12}}{C_{21} + C_{23}}$$

For all event types the Robinson and Smyth exact test (Robinson & Smyth, 2008) was performed in R 3.1.3 with the Benjamini & Hochberg multiple testing correction method being applied to determine if there was DAS.

2.2.4.3 SUPPA with diffSplice

SUPPA version (2.1.0) was used to generate the different AS events from AtRTD and the PSI of events was calculated using a similar command to that in section 2.2.4.1, with the main difference being SUPPA version (2.1.0) is run using python 3.4 (there is limited backwards compatibility between python 2 and python 3, hence an error will result if earlier version of python is used to run SUPPA 2.1.0). The other difference is that the mock and infected PSI were determined separately, for example

```
suppa.py psiPerEvent – GeneratedEvents_RI.ioe –e
Sailfish_TPM_Mock.txt -o SUPPA_Sailfish_TPM_RI_Mock
```

and

```
suppa.py psiPerEvent – GeneratedEvents_RI.ioe –e
Sailfish_TPM_Inf.txt -o SUPPA_Sailfish_TPM_RI_Inf
```

Differential splicing events were calculated using the diffSplice function using the empirical method; an example of which is:

```
python3 suppa.py diffSplice -m empirical -p
SUPPA_Sailfish_TPM_RI_Mock.psi
SUPPA_Sailfish_TPM_RI_Inf.psi -e Sailfish_TPM_RI_Mock.txt
SUPPA_Sailfish_TPM_RI_Inf.txt -i GeneratedEventsTest_RI.ioe
-o SUPPA_Sailfish_TPM_RI_emperical
```

This was repeated for all events and quantifiers; a generic usage of the program is given below:

```
suppa.py diffSplice --method <empirical> --ioe <ioe-file> --
psi <Cond1.psi> <Cond2.psi> --expression-file
<Cond1_expression-file> <Cond2_expression-file> -o <output-
file>
```

2.2.5 Splice junction level algorithms

Junctions files from Tophat2 or STAR were used in Limma, run in R version 3.1.3 to determine statistically significant differential usage of SJs utilising the linear modelling approach implemented by lmFit and the empirical Bayes statistics implemented by eBayes; SJ that had less than 3 reads in all treatments were filtered out and the "voomWithQualityWeights" option was used to provide a more powerful analysis as recommended by Liu *et al* 2015. An example code can be found in section 7.3.

2.2.6 Gene ontology enrichment

GO enrichment analysis was performed using BINGO 3.0.3 (Maere, Heymans, & Kuiper, 2005). The biological GO terms used for Arabidopsis were downloaded from http://www.arabidopsis.org on 29 September 2015. A hypergeometric test for over-represented / under-represented GO terms with an FDR of 0.05 using the Benjamini Hochberg multiple testing correction was utilised.

2.2.7 Predicting changes in proteins

To investigate if AS events affected the domains of the proteins produced, the nucleotide sequence was obtained from TAIR (http://www.arabidopsis.org), and ran through the NCBI Conserved Domain Database (http://www.ncbi.nlm.nih.gov/cdd)(Marchler-Bauer *et al*., 2015). The differences in protein domains between isoforms were visually compared.

2.2.8 Determining enriched motifs

DREME on CyVerse was utilised to determine which motifs were enriched compared to the TAIR10 RI set determined by Mao *et al* 2014 (Bailey, 2011; Goff *et al*., 2011; Mao, Raj Kumar, Guo, Zhang, & Liang, 2014), with a motif range of 3 to 12 nt unless otherwise stated in the relevant results sections.

3 Chapter 3: Differential Alternative Splicing Occurs in Arabidopsis Leaves in Response to Infection with The Necrotrophic Fungal Pathogen B. cinerea

3.1 Introduction

It has previously been shown that *Arabidopsis* leaf tissue undergoes genome-wide transcriptional reprogramming following infection with the fungal pathogen *Botrytis cinerea*, with approximately one third of the Arabidopsis genome changing in expression during the first 48 hours post infection (hpi) (Windram *et al*., 2012). This in-depth time series was carried out using microarrays that could not detect gene transcripts, and hence it was unable to detect differential alternative splicing (DAS). A dataset created using RNA-Seq to investigate global transcriptional change in Arabidopsis leaves in response to *B. cinerea* infection by Cooke (2013) exists. This data set can be used to investigate DAS. However it only has three time points and only one biological replicate (Cooke, 2013).

In this dataset Cooke (2013) identified genes that show DAS using RNA-Seq gene expression profiling of Arabidopsis leaves at 14, 24, and 34 hpi inoculated with *B. cinerea* or mock-inoculated. These time points were selected as they represent key infection stages; host penetration by the fungus and initial increase in fungal mass, a lag phase in fungal growth coinciding with a large change in host gene expression, and, finally, the further increase in fungal mass as the lesion spreads throughout the leaf tissue (Cooke, 2013; Windram *et al*., 2012). Cooke (2013) used three different methods to identify DAS genes; a transcript level approach using known gene models from TAIR10 and Marquez *et al* (2012), an event level approach using the program MISO (Katz, Wang, Airoldi, & Burge, 2010) and a splice junction level approach that identified incompatible splice junctions (i.e. splicing junctions that are mutually exclusive) (Cooke, 2013; Marquez, Brown, Simpson, Barta, & Kalyna, 2012)}. The analysis identified DAS in response to infection as well as over time (the

experimental samples, 14, 24 and 34 hpi, span a period of 20 hours). Using all three of the above methods 4% of intron-containing genes (924 out of 22302) in *Arabidopsis* exhibited DAS due to *B. cinerea* infection; 637 of these were also differentially expressed with 179 up-regulated across all time points and 237 down-regulated across all time points, 78 genes were up-regulated and 123 were down-regulated at two or more consecutive time points, with the remaining 20 genes being transient in nature in response to infection. In addition 171 genes were found to DAS due to time alone with no additional response to infection (Cooke, 2013). This indicates that there is a relatively large amount of DAS occurring in response to *B. cinerea*, which appears not necessarily correlated with differential gene expression. This dataset was created in the early days of RNA-Seq when investigating DAS in plants was a relatively new field. Here we will utilise newer tools and techniques to explore the DAS genes Cooke (2013) identified, validate a subset of genes that we are confident undergo DAS in response to *B. cinerea* that can be used as marker genes in further research, as well as investigating whether some of these genes play a role in plant defence.

3.2 Aims and Objectives

The overarching aim of this chapter is to determine if DAS occurs in Arabidopsis in response to *B. cinerea* and to begin to understand what roles it could potentially play in the plant defence response. In order to achieve this the following research objectives will be met:

1. Validate and evaluate the robustness of the Cooke (2013) dataset by
    a. Comparing differentially expressed genes (DEG) in the dataset to known Arabidopsis marker genes for *B. cinerea* infection
    b. Compare the dynamics of the DEG in Cooke (2013) with those from the in depth microarray time series of Arabidopsis leaves infected with *B. cinerea* carried out by Windram *et al* 2012(Cooke, 2013; Windram *et al*., 2012)
    c. Compare genes deemed to undergo DAS in response to *B. cinerea* infection by Cooke (2013) with other datasets investigating splicing
2. Expand on the knowledge gained by Cooke (2013) by utilising the set of genes deemed to be DAS in response to *B. cinerea* to:
    a. Identify potential *B. cinerea* responsive DAS marker genes
    b. Link DAS gene/s to a defence phenotype and determine if this is pathogen specific
    c. Determine if RI events that are DAS in response to *B. cinerea* are enriched for DNA binding motifs

3.3 Results

The RNA-Seq dataset provided by Cooke (2013) consisted of one sample per time point for mock-inoculated leaves and two technical replicates per sample for *B. cinerea* infected leaves that were processed to give paired-end 76 bp reads. To determine DEG, Cooke (2013) aligned the reads using Bowtie, counted the number of reads that aligned to each gene in TAIR 10, and the count data was processed using a generalised linear model (GLM) to determine genes differentially expressed (DE) between either time or treatment (this also included those that were DE in both).

3.3.1 Validating the Cooke dataset

To help determine the accuracy and reliability of the Cooke (2013) dataset those genes that are identified as differentially expressed and DAS were investigated.

3.3.1.1 Validation of differentially expressed genes

The DEGs included some well-known *B. cinerea*-responsive genes such as Phytoalexin Deficient 3 (*PAD3*), Plant Defensin 1.2 (*PDF1.2*) and WRKY DNA-Binding Protein 33 (*WRKY33*) (Ferrari *et al*., 2007; Ferrari *et al*., 2003; Zheng *et al*.,2006), which can be utilised as marker genes. The progress of infection in the RNA-Seq experiment was remarkably similar to that of Windram *et al* 2012 (Windram *et al*., 2012) as seen by the kinetics of *WRKY33* expression in both data sets (Figure 13A, B). The peak of expression is seen at the same time in both experiments, indicating that Cooke obtained reproducible infection conditions. *PAD3* and *PDF1.2* were not included on the Windram microarray, however the expression profile of *PAD3* and *PDF1.2* (Figure 13C, D) follows the same pattern as previously published data (Ferrari *et al*., 2003; 2007).

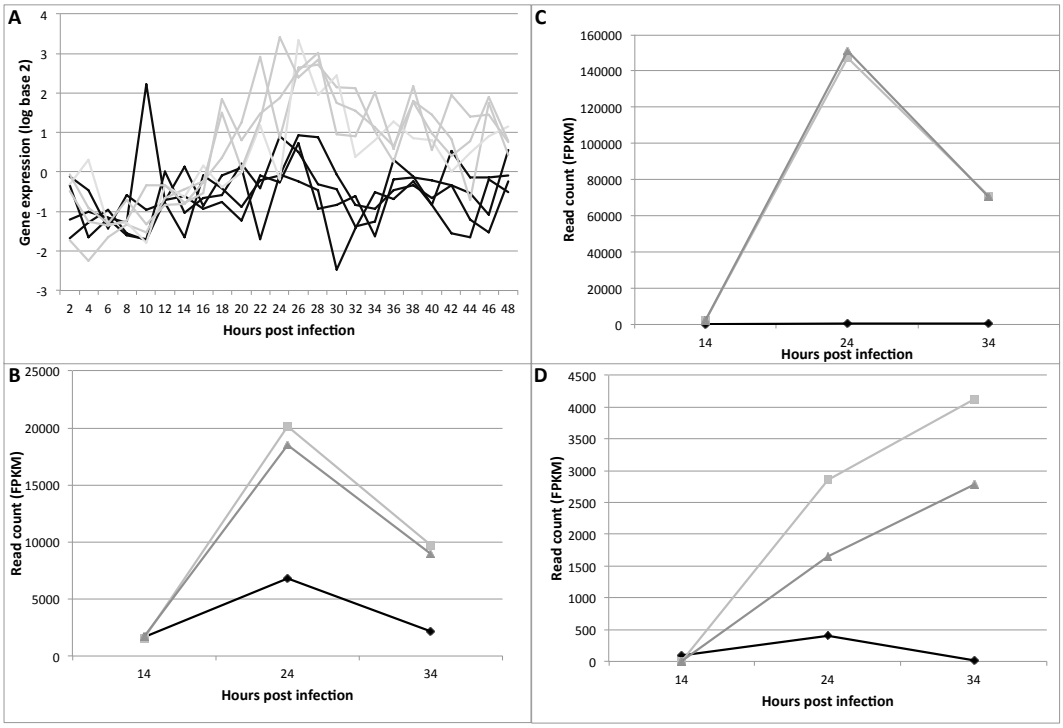


Figure 13. Expression of key marker genes induced in response to *B. cinerea* infection. *WRKY33* encodes a transcription factor and has the highest absolute gene expression, followed by *PAD3* which encodes an enzyme, with *PDF1.2*, which encodes a peptide, having the lowest gene expression. A) Microarray data across the time-course of *B. cinerea* infection showing expression of WRKY33. B-D RNA-Seq data showing expression of B) WRKY33, C) PAD3 and D) PDF1.2. Grey lines represent *B. cinerea* infected samples and black lines represent mock samples.

The absolute expression of these three important defence genes can be inferred from their read coverage, which was normalised for library and gene size (Cooke, 2013).The *WRKY33* gene which encodes a transcription factor has the highest overall expression followed by *PAD3* which encodes an enzyme, and the gene that encodes a peptide (*PDF1.2*) having the lowest expression level of the three (Figure 13), this is interesting because transcription factors are often thought to be expressed at low levels. Evaluating these three marker genes highlights that the Cooke (2013) DEG dataset contains genes covering a large dynamic range of expression, including genes known to be DE in response to *B. cinerea*.

Although some of the DEGs did not have probes on CATMA v3 microarrays, Cooke (2013) states that the RNA-Seq profiling has identified a significantly larger number

of genes than the full time series analysis (Cooke, 2013; Windram *et al*., 2012). Of the 12940 DEGs, 12105 of these were present on the CATMA v3 microarrays, and thus could have been detected in the Windram *et al* 2012 study (Windram *et al*., 2012). Over half of these were identified as differentially expressed in response to infection by both Windram *et al* (2012) and Cooke (2013). The RNA-Seq analysis, however, found that over 5000 DEGs were uniquely detected by the RNA-Seq experiment (Figure 14) (Cooke, 2013; Windram *et al*., 2012).

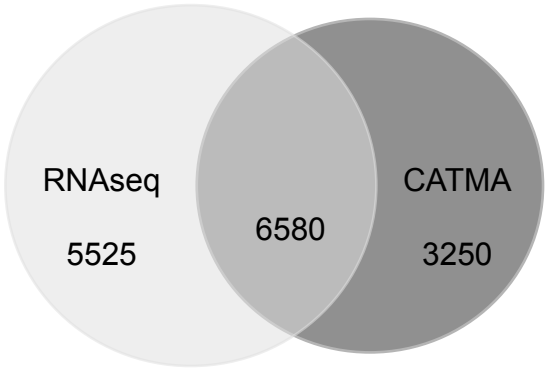


Figure 14. ***RNA-Seq detects more differentially expressed genes when compared to CATMA arrays.*** **Of the 12940 genes differentially expressed due to infection (FDR of 1%) 12105 of these have CATMA v3 probes. Windram *et al* (2012) found 9838 genes (9830 with gene models in TAIR10) that are differentially expressed due to infection using CATMA v3 probes, while Cooke (2013) detected 6580 of these using RNA-Seq. 5525 differentially expressed genes were only detected using RNA-Seq. Figure adapted from Cooke (2013).**

To investigate the likelihood of these genes are involved in the stress response and thus likely to be true positives Gene Ontology (GO) enrichment analysis was performed using BINGO 3.0.3 (Maere, *et al*., 2005) with an updated GO annotation file (September 2015). The DEGs identified by both studies and the DEGs identified by RNA-Seq alone show a very similar pattern of GO term overrepresentation with enrichment of expected terms such as “response to stress”, “response to biotic stimulus” and “response to fungus” indicating that those additional genes identified by the RNA-Seq experiment potentially play a role in the defence response (Figure 15).

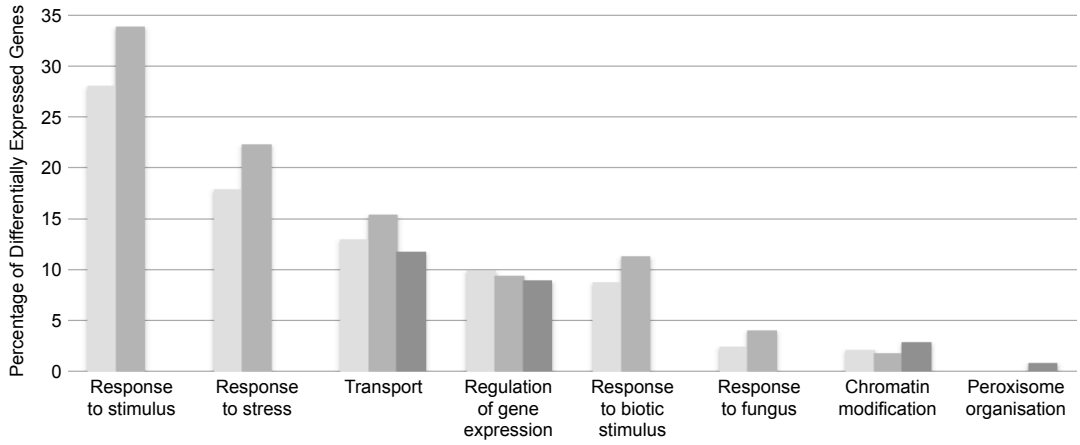


Figure 15. Functional analysis (GO annotation) indicates biologically interesting genes are detected in the RNA-Seq data set. The number of DEG over-represented for GO terms in the RNA-Seq dataset only, both datasets, and the CATMA microarray dataset only, are represented by light grey, medium grey, and dark grey bars respectively.

By comparing the normalised read count between those genes detected to be DE only in the RNA-Seq data, and those deemed to be DE in both the RNA-Seq and CATMA microarray data, we can infer that the absolute change in expression is lower for those genes that are only detected in the RNA-Seq dataset at all time points (one way ANOVA $p < 0.01$). This potentially indicates that the RNA-Seq method is more sensitive at detecting DEG than the CATMA microarrays (Figure 16).

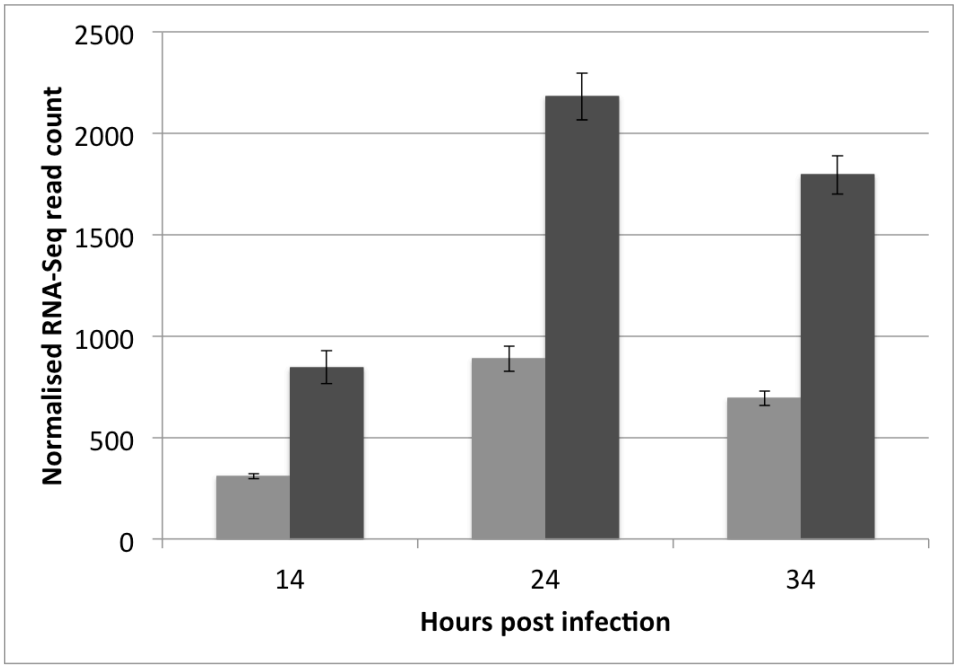


Figure 16. RNA-Seq is more sensitive than CATMA microarrays. DEG that were detected in both RNA-Seq and CATMA arrays have a significant higher gene expression than those detected by RNA-Seq only at all time points (Student T-test, $p<0.01$); this is inferred from the normalised RNA-Seq read count. Light grey and dark grey bars represent genes detected to be DE in only the RNA-Seq dataset, and in both RNA-Seq dataset and the CATMA microarray dataset, respectively. Error bars represent SD.

Why RNA-Seq does not detect the 3250 DEGs identified only in the CATMA array study is not clear. It is possible that this may be due to transient differential expression of these genes as the CATMA array experiment was carried out over a 48-hour period with samples collected every 2 hours, or genes that are differentially expressed after 34 hours.

This analysis of the RNA-Seq dataset indicates that the *B. cinerea* infection is proceeding as expected and with regards to DEGs. It appears similar to a previously published dataset on Arabidopsis leaves inoculated with *B. cinerea* with expected key marker genes being shown to be DE. This indicates that we can have a fairly high level of confidence in the accuracy and reliability of the data. In addition compared to CATMA microarrays, RNA-Seq appears to be a more sensitive technique that can identify a greater number of truly DEG.

3.3.1.2 Investigating DAS genes

In the Cooke (2013) dataset 4% of Arabidopsis genes undergo DAS in response to *B. cinerea* infection. This is in line with previous studies investigating DAS in response to *Pseudomonas syringae* pathovar *tomato* (*Pst*) and salt stress in Arabidopsis, where 4% and 10% of genes were differentially alternatively spliced respectively (Cooke, 2013; Ding *et al*., 2014; Howard *et al*., 2013).

When analysing AS events, they can be grouped into 4 main categories: retained introns (RI), skipped exons (including alternative first and alternative last exons) (SE), alternative 3' splice sites (A3SS) and alternative 5' splice sites (A5SS). A single gene could have multiple isoforms and events. However Cooke (2013) analysis was carried out in a pair-wise manner, hence each differentially alternatively spliced event will fall into one of these categories. Of those genes Cooke (2013) detected to be DAS at the event level using both TAIR10 and Marquez *et al* 2012 gene models, 694 differentially spliced events occurred due to infection (or a combination of time and infection) from 582 unique genes, with 118 DAS events originating from 109 unique genes being identified in response to time alone (i.e. they were not influenced by infection) (Cooke, 2013; Marquez *et al*., 2012). Mining the dataset provided by Cooke (2013) we ascertained that although the number of DAS events is greater in response to infection than time, the frequency of each type of event is very similar, with RI events being the most prevalent event type in response to infection and time, with 77% and 71% of total events respectively (Figure 17).

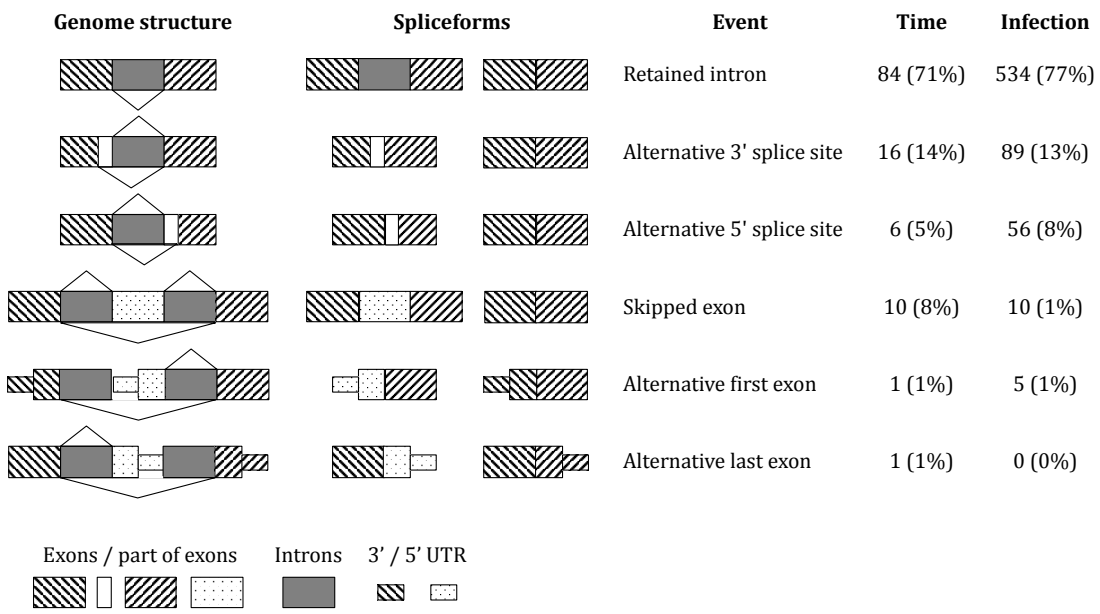


Figure 17. The proportion of event type of DAS events in response to *B. cinerea* infection and due to time alone are roughly the same. The first column illustrates the intron-exon structure of the AS event, followed the possible spliceforms arising from such events and its description. There were 118 and 694 DAS events due to time and infection, respectively, with the proportion of each event type being roughly the same in each case.

The prevalence of RI events compared to other types of DAS events agrees with previous AS studies (Iida *et al*., 2009; Marquez *et al*., 2012; McGuire, Pearson, & Neafsey, 2008; B.-B. Wang & Brendel, 2004a) including those concerned with detecting DAS genes due to stress (Ding *et al*., 2014; Howard *et al*., 2013). The AS studies have shown the proportion of RI events is between 39 and 56% of the total AS events (Iida *et al*., 2009; Marquez *et al*., 2012; McGuire *et al*., 2008; B.-B. Wang & Brendel, 2004a), which is considerably less than the proportion of RI events detected. However, these studies were investigating AS under normal growth conditions, rather than investigating DAS genes between plants grown under normal and stress conditions. Recent studies by Howard *et al* (2013) and Ding *et al* (2014) found that the number of differentially expressed RI events between control plants and those under stress conditions were 70.6% and 67% respectively, which is in a similar range to these findings (77%) (Ding *et al*., 2014; Howard *et al*., 2013). However, a RI rate of 70% is considered high and could potentially represent partially spliced transcripts rather that a true differentially alternatively spliced event.

When performing RNA-Seq a small proportion of partially processed transcripts that have not had their introns spliced out can be sequenced. To prevent these transcripts being counted as RI events Cooke (2013) disregarded any event where the gene model did not have ten or more reads assigned to it. To investigate this further, we investigated whether transcripts with an infection-responsive RI event show full splicing of other introns in the same gene. Of those genes that contain RI events, 471 were in genes that contained more than one intron. Only seventeen of these genes had all of the introns retained, with 454 genes showing splicing of at least one other intron. This suggests that effective RNA processing has occurred and the RI events are a real biological phenomenon.

To analyse the function of the genes showing infection-responsive DAS, GO enrichment analysis was performed using BINGO 3.0.3 (Maere *et al*., 2005) and the representative subsets determined using the REVIGO package (a clustering algorithm based on semantic similarity measurements to provide representative subgroups or clusters of GO terms) (Supek *et al*., 2011). Genes DAS in response to time alone were enriched for different subsets of GO terms than those DAS in response to *B. cinerea* (Figure 18). Those genes DAS in response to *B. cinerea* infection were over-represented for a variety of biological processes, including "response to other organism" as would be expected if these DAS genes are involved in plant defence (Figure 18A), crucially, the genes showing alternative splicing in response to time alone appear to have completely different functions (Figure 18B).

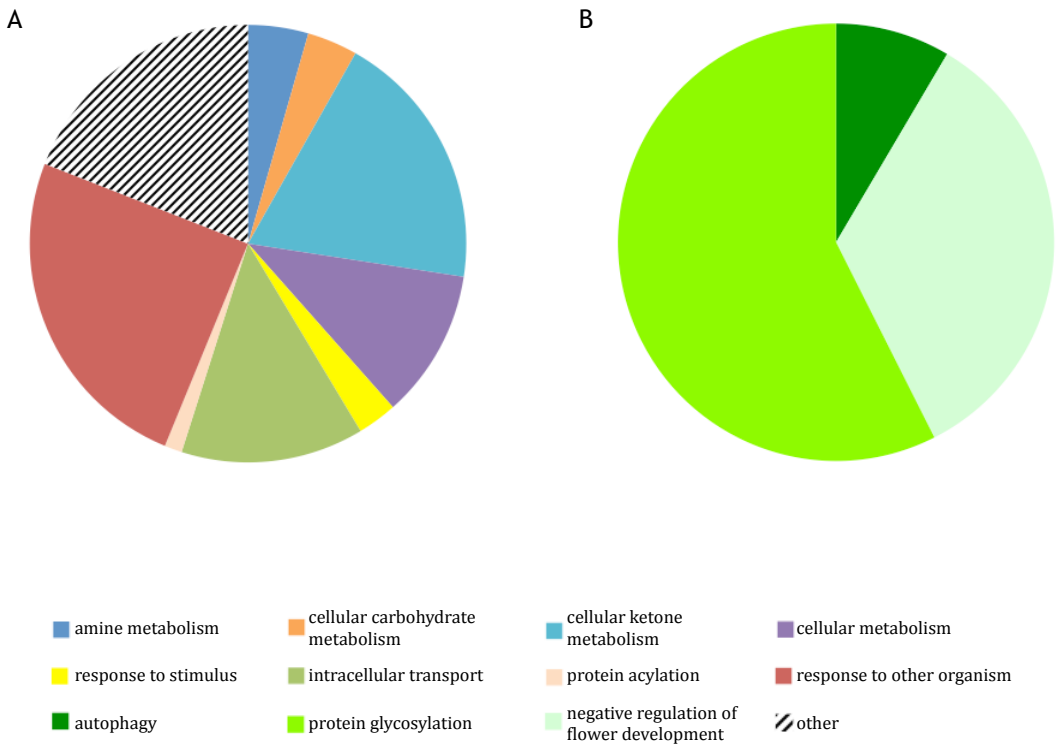


Figure 18. Biologically significant DAS genes due to *B. cinerea* infection are detected using RNA-Seq. Genes that are DAS in response to *B. cinerea* infection are over-represented by different GO terms compared to those DAS due to time alone (A and B respectively). The genes that are differentially spliced in response to *B. cinerea* infection are enriched for defence related GO terms such as "response to stimulus" and "response to other organism". This diagram represents data from Cooke (2013).

Overall the comparisons made between the Cooke (2013) dataset and previously published works combined with GO term functional analysis indicates that it is of sufficient quality and relevance to use to select individual genes to use as potential DAS marker genes and to further explore the role of DAS in response to *B. cinerea* infection.

3.3.2 *B. cinerea*-mediated DAS genes cover a broad range of functions

Comparing the Cooke (2013) dataset of genes DAS in response to *B. cinerea* with publically available datasets of genes DAS in response to other stresses could help to elucidate if DAS is a general stress response. Mining the datasets of genes deemed to be DAS in response to *Pst* and salt stress shows that 24 and 244 genes respectively were also detected to undergo DAS in response to *B. cinerea* infection,

with 12 of these genes undergoing DAS in response to all three stresses (Ding *et al*., 2014; Howard *et al*., 2013) (Figure 19).

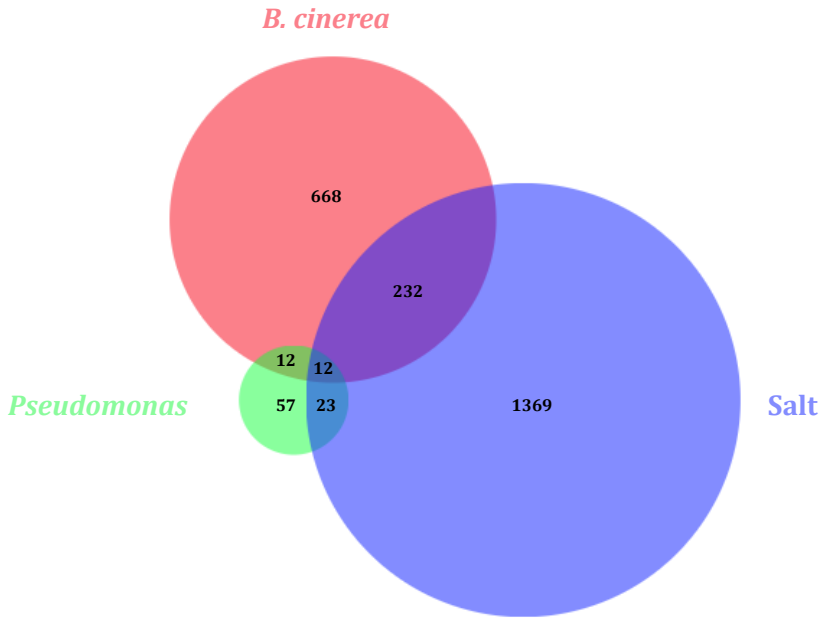


Figure 19. Genes are DAS in Arabidopsis in response to *B. cinerea*, *Pst* and salt stress. Each of the stresses produce unique DAS genes and genes that are DAS in more than one stress. Red, green and blue circles represent *B. cinerea*, *Pst* and Salt stress respectively. Venn Diagram produced using Biovenn (Hulsen, de Vlieg, & Alkema, 2008).

B. cinerea and *Pst* represents biotic stresses, whereas salt stress represents an abiotic stress, to determine if genes that overlap between these different types of stress could have different functions, GO enrichment analysis was performed using BINGO 3.0.3 (Maere *et al*., 2005). Some GO terms are only enriched in genes that are DAS in response to *B. cinerea* and not the other stresses, including those specifically associated with fungal pathogens, for example "response to chitin" and "response to ethylene stimulus", indicating that these genes may be DAS in a stress-specific manner (Figure 20). Whilst genes that are DAS under all three stresses are enriched for terms that indicate a more general role in plant defence response, for example the GO terms "response to oxidative stress" was enriched for genes DAS under *B. cinerea*, *Pst* and salt stress, as were GO terms associated with the JA signalling pathway such as "jasmonic acid mediated signalling pathway" (Figure 20). It has long been known that the plant defence response involves the JA pathway and recently

Valenzuela *et al* (2016) have shown that this signalling pathway is also triggered in response to salt stress, which could explain why genes DAS in response to both of these types of stresses are enriched for this term.

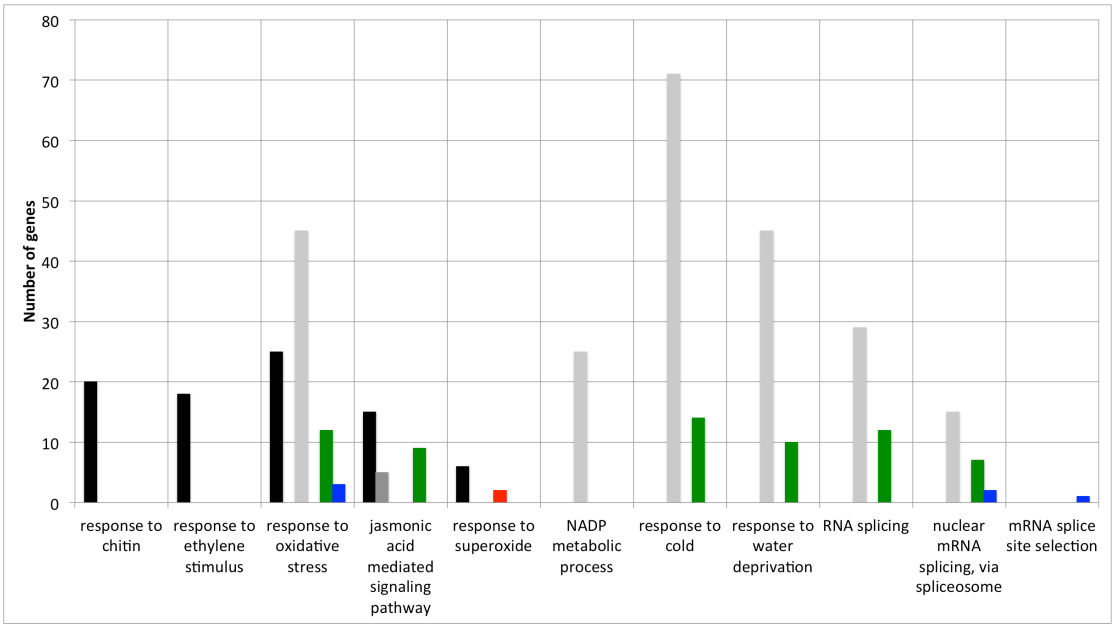


Figure 20. GO term enrichment indicates that the function of genes undergoing DAS in response to stress can be stress specific, or part of a general stress response. Black, dark grey and light grey show the number of statistically significant (p<0.01) genes over-represented for GO terms in samples treated with *B. cinerea*, *Pst*, NaCl. The red, green and blue bars represent the number of genes whose related GO terms that are over-represented in more than one stress in the following combinations, red represents *B. cinerea* and *Pst*, green represents *B. cinerea* and NaCl, and blue represents all three stresses.

Genes DAS in response to only the biotic stresses were enriched for the GO term "response to superoxide", this term was not enriched in genes DAS in response to salt stress indicating that some DAS may be stress type-specific. Further evidence for this is the enrichment of the GO term "NADP metabolic processes" that was uniquely found in genes DAS in response to salt stress (Figure 20). Interestingly some GO terms were enriched in genes that were DAS in response to *B. cinerea* and salt stress but not *Pst*, for example GO terms associated with abiotic stresses such as "response to cold" and "response to water deprivation" possibly indicating a dual function for these genes (Figure 20). GO terms associated with splicing were also enriched in genes DAS in response to all three of the stresses for example "nuclear mRNA splicing, via spliceosome" and "mRNA splice site selection" indicating that splicing

plays a functional role in response to stress, although the GO term "RNA splicing" was not enriched in *Pst* and the number of genes associated with these GO term are quite low (Figure 20). This GO enrichment analysis indicates that DAS plays a role in both biotic and abiotic stresses, with some genes being DAS as part of the general stress response and some that are DAS in a more defence-specific manner.

3.3.3 Validation of a subset of genes shown to be differentially alternatively spliced in response to *B. cinerea*

Eight genes, chosen because they undergo a variety of *B. cinerea* mediated DAS events, and which cover a range of proportional changes in isoforms between mock and infected samples in the Cooke (2013) dataset were chosen to be investigated further. Seventeen isoforms, from 8 genes, were tested in three new biological samples using an established high resolution reverse transcription polymerase chain reaction HR RT-PCR (Simpson *et al*., 2008)[1]. For twelve of these, the RQ-RT PCR demonstrated statistically significant DAS in response to *B. cinerea* infection . For the gene *AT4G39270*, the RNA-Seq data indicated that the largest change in isoforms occurs between the fully spliced version (*AT4G39270.1*) and the isoform where all the introns are retained (*AT4G39270_ID4*). Usually, using the HR RT-PCR method the events where all introns are retained in a transcript are not scored because this may be due to DNA contamination. Re-scoring of the transcript that has all three introns retained shows that this transcript (AT4G39270_ID4) is reduced in response to *B. cinerea* infection, consistant with the RNA-Seq data. However this was only carried out for one biological sample.

[1] Experimental procedures were carried out during my MSc project (Foster, 2012); the analysis was carried out during my PhD.

Table 6. Differential alternative splicing events are validated using High resolution RT-PCR. 17 isoforms that are deemed to be DAS in response to *B. cinerea* infection (column three) originating from eight genes (column one) with a variety of DAS events (columnfour) underwent validation using high resolution RT-PCR. For each of the isoforms, except for AT4G39270_ID4 (denoted with *), three biological replicates per treatment (mock, 24 hpi and 34 hpi) were analysed. Student T-test was performed to determine if statistically significantly DAS in response to infection occurred at the 5% level, with the null hypothesises being that there is no difference between mock and infected samples. AT4G39270_ID4 appeared to be DAS in response to infection but only one biological sample was scored.

Gene	Brief Description	Isoform	Event	24 hpi	34 hpi
AT1G28330	Dormancy-associated protein (AtDMR1)	AT1G28330.1	Fully Spliced	No	Yes
AT1G28330		AT1G28330.2	SE (Exon 3)	No	Yes
AT1G74590	Encodes tau class glutathione transferase (AtGSTU10)	AT1G74590_ID1	Fully Spliced	Yes	Yes
AT1G74590		AT1G74590_ID2	RI (Intron 1)	Yes	Yes
AT3G02300	Regulator of chromosome condensation family protein.	AT3G02300_ID2	Fully Spliced	No	Yes
AT3G02300		AT3G02300_ID3	A5SS (Exon 6)	No	Yes
AT3G59350	Protein kinase	AT3G59350.1	Fully Spliced	No	No
AT3G59350		AT3G59350.2	AFE	No	No
AT4G20480	Putative endonuclease of glycosyl hydrolase	AT4G20480_ID1	Fully Spliced	Yes	Yes
AT4G20480		AT4G20480_ID2	RI(Intron 3)	Yes	Yes
AT4G39270	Leucine-rich repeat protein kinase	AT4G39270.1	Fully Spliced	No	No
AT4G39270		AT4G39270.2	RI (Intron 3)	No	No
AT4G39270		AT4G39270_ID4	RI (Introns 1, 2, & 3)	Yes*	Yes*
AT5G16800	Acyl-CoA N-acyltransferase	AT5G16800.1	Fully Spliced	Yes	Yes
AT5G16800		AT5G16800.2	RI (Intron 4)	Yes	Yes
AT5G48657	Defence related protein	AT5G48657.1	RI (Intron1)	Yes	Yes
AT5G48657		AT5G48657.2	Fully Spliced	Yes	Yes

Because only one repeat biological sample for *AT4G39270* was scored, further validation of the *B. cinerea* mediated DAS of the gene *AT4G39270* was required, this was carried out using PCR. The Cooke (2013) RNA-Seq data indicates that *AT4G39270* undergoes one of the largest proportional changes in isoforms with the amount of the fully retained intron isoform increasing in response to *B. cinerea* infection. To investigate this, primers were designed that spanned all three introns

and PCR combined with gel electrophoresis was carried out in three new biological samples at 24 and 34 hpi[2], with the results validating the RNA-Seq data (Figure 21).

Visual inspection of the agarose gel shows that in mock-inoculated samples both isoforms are present, whereas in the *B. cinerea* infected samples only the fully spliced transcript (*AT4G3970.1*) is present (Figure 21C). This results in a validation rate for the Cooke (2013) RNA-Seq DAS data set of 76.5% (13 out of 17 isoforms validated), which is in line with that of other RNA-Seq studies detecting alternative splicing and DAS in response to stress in plants (Filichkin *et al*., 2010; Howard *et al*., 2013).

[2] Experimental procedures were carried out during my MSc project (Foster, 2012); the analysis of band intensities was carried out during my PhD.

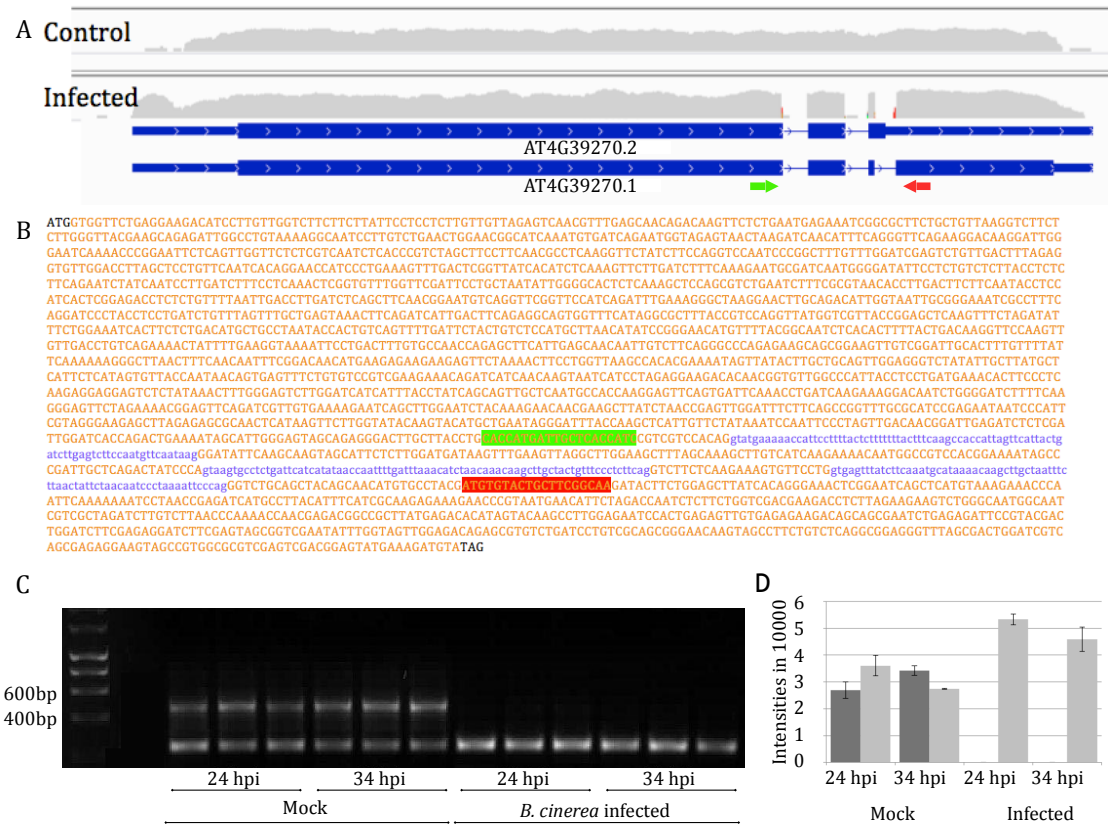


Figure 21. AT4G39270 is DAS in response to *B. cinerea* infection. A. RNA-Seq data by Cooke (2013) indicates that the introns are retained in uninfected samples and spliced out in response to infection with *B. cinerea*. Blue bars are predicted transcripts from TAIR10 database, height of grey bars indicates number of times each base pair was sequenced. Green and red arrows indicate forward and reverse primers respectively, diagram adapted from Cooke (2013) (Cooke, 2013). B. Primers were designed to produce PCR products that contained all three introns, green and red highlighted areas indicate forward and reverse primers respectively. The blue letters represent introns, thus by counting the number of nucleotides between the two primers we obtain the expected sizes for the retained intron (AT4G39270_ID4) and the fully spliced (AT4G39270.1) transcripts are 475bp and 233bp respectively. C-D. PCR results validate *B. cinerea* mediated DAS of AT4G39270. C. Two splice transcripts are seen in mock samples at 24 hpi, and 34 hpi. In samples infected with *B. cinerea* only the retained intron splice transcript is visible on the gel. Diagram adapted from Foster (2012). D. Analysing the gel image in ImageJ to determine the intensities of the bands indicates that the mock samples contain approximately half of each splice variant compared to infected samples where the retained intron spliceform (dark grey bars) are not seen. Light grey bars represent spliced spliceform. Error bars are the SEM of 3 biological replicates.

To obtain a semi-quantitative measure of the amount of *AT4G39270* isoforms present the gel electrophoresis band intensities were analysed in ImageJ; with roughly the same amount of the fully spliced isoform (*AT4G39270.1*) and the isoform with all introns retained (*AT4G39270_ID4*) being present in the mock samples

(Figure 21D). There was no visible band relating to *AT4G39270_ID4* in infected samples so a band intensity could not be determined. However the intensity of the infected samples is not equal to the combined intensities of *AT4G39270.1* and *AT4G39270_ID4* found in the mock, potentially indicating that there is a small amount of the *AT4G39270_ID4* present in infected samples but the amount is too low to be determined using standard PCR. There was no evidence found of isoform *AT4G39270.2*. However due to the shortness of read length and the depth of sequencing this RNA-Seq dataset cannot differentiate between *AT4G39270.1* and *AT4G39270.2* because not many reads start in the right position to be able to distinguish between the two isoforms. Both *AT4G39270.1* and *Atg39270.2* are predicted to encode fully functional proteins and HR RT-PCR shows that these two spliceforms are not DASed in response to *B. cinerea* infection at either 24 hpi or 34 hpi (Students T-test p=0.09 and p=0.14 for 24 and 34 hpi respectively), with *AT4G39270.1* being the predominant spliceform (on average 98% of transcripts that encode functional protein are AT439270.1 with the other 2% being AT4G39270.2) (Figure 22).

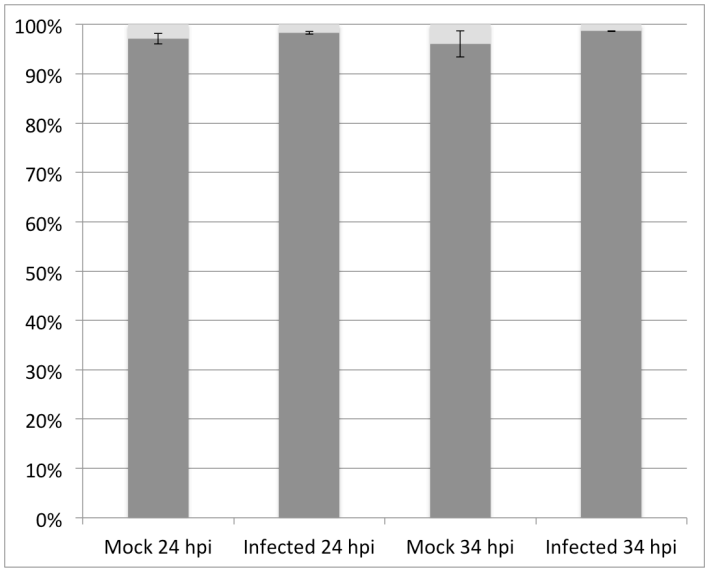


Figure 22. The relative amount of AT4G39270.1 to AT4G39270.2 is not affected by *B. cinerea* infection. Comparing the amount of the transcripts that are predicted to encode the functional protein shows that the relative amount of these two spliceforms does not change in response to infection (Students T-test p>0.05 at both time points). On average (mean) At4g39270.1 accounts for 98% of transcripts that are predicted to encode functional protein. Dark grey bars represent the proportion of AT4G39270.1 and light grey bars represent the proportion of AT4G39270.2

To further investigate the RI spliceform compared to those spliceforms that are predicted to produce proteins, quantitative Real-time quantitative reverse transcription polymerase chain reaction (RQ-RT PCR) was utilised because it is a quantitative method with a larger dynamic range. When using RQ-RT PCR primers producing shorter amplicon sizes are typically used because they generally have a higher efficiency, thus creating RQ-RT PCR primers that span the entire gene could be problematic. This was the case when utilising the above primers for RQ-RT PCR with the efficiencies being very low, thus new primers were designed. Investigating the gene sequence of *AT4G39270_ID4* showed that the retention of intron one causes a frame shift resulting in a PTC occurring just after the first intron (Figure 23A). Thus, to further investigate, *AT4G39270_ID4* primers that investigated the retention of the first intron were used because if this intron is retained then it is immaterial if the other introns are retained or not (because the transcript will terminate before the second intron). This enables primers that produce relatively short amplicons to be designed. Two forward primers were used to investigate the retention of intron one of *AT4G39270*, one that spanned the intron and one that is inside the intron, with the same reverse primer used for both, resulting in amplicon sizes of 212 bp and 174 bp for the spliced and retained intron respectively (Figure 23B).

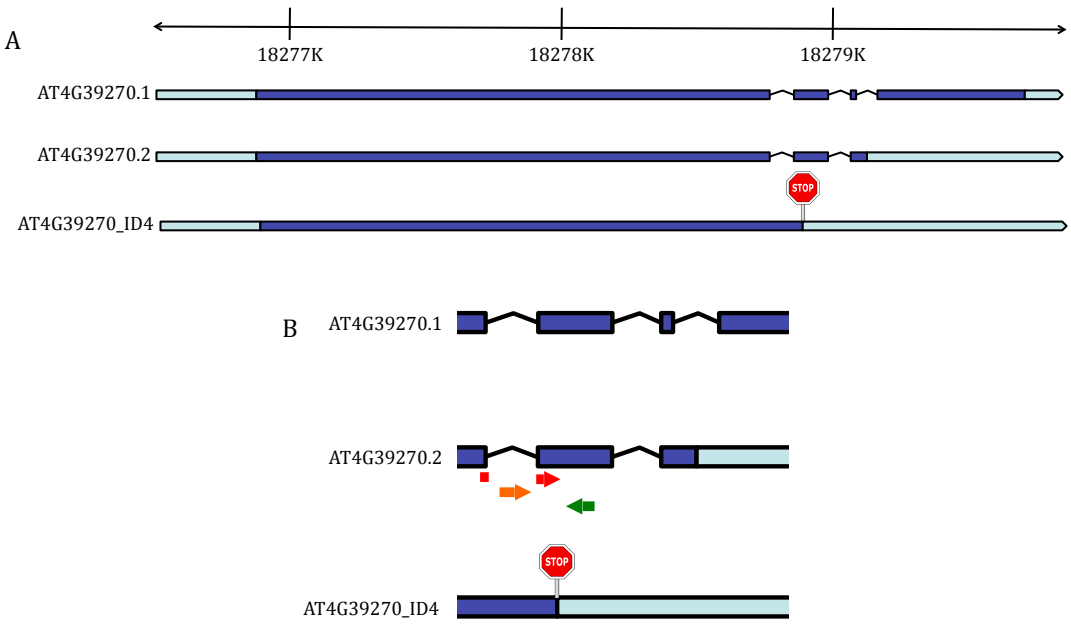


Figure 23. Gene models for AT4G39270. A. *AT4G39270.1* and *AT4G39270.2* have three and two introns spliced respectively, AT4G39270_ID4 has all three introns retained, and the retention of the first intron results in a premature termination codon, represented by stop sign. B. To investigate the retention of intron one, two forward primers were designed, one that spanned the intron (red) and one that is located within the intron (orange). The same reverse primer (green) was paired with each forward primer.

Inspection of the quantification cycle (Cq) values obtained from RQ-RT PCR indicated that the RI spliceform was present in all samples. A standard curve reaction was performed and the Cq values plotted to obtain the slope of the graph with the primer efficiencies and amplification factor calculated using the thermoscientific online RQ-RT PCR efficiency calculator (http://www.thermoscientificbio.com/webtools/RQ-RT PCRefficiency/). This indicated that the efficiencies of the primers were 75.4% and 73.5% for the RI and spliced intron (SI) amplicon respectively, with the amplicon factor being 1.75 and 1.74. Although these efficiencies are lower than ideal they are both of similar values, with a limited area of DNA to design the primers in due to the features we wanted to include (primers across and within an intron), it was decided to use these primers and apply the Pfaffl variation of the ΔΔcq method to analysis the RQ-RT PCR data (Pfaffl, 2001). The Pfaffl method was used because the efficiency of both primer pairs was lower than 90%. The ΔΔcq method assumes a two-fold increase in PCR

product every cycle but this would not be the case with lower efficiency primers, thus a 1.74 correction was used based on the values of the amplification factor calculated using the thermoscientific online RQ-RT PCR efficiency calculator (Pfaffl, 2001).

Different biological samples may contain different amounts of RNA. To account for this the relative amount of the SI compared to the RI was calculated. There is between four and six times more of the SI amplicon relative to the RI amplicon in mock samples, with similar values seen in *B. cinerea* infected samples at 14 hpi (five-fold increase of SI relative to RI). At 24 hpi the amount of SI is significantly more with 15 times the amount relative to the RI. The amount of SI relative to the RI continues to increase over the course of infection, with 30 and 45 times the amount of SI relative to RI detected at 34 and 48 hpi respectively (Figure 24). These time course results indicate that in uninfected leaves there is approximately five times the amount of the fully spliced transcript compared to the transcript where the first intron is retained, and this amount is unchanged in early stage of infection. At 24 hpi when the lag phase of fungal growth occurs, the relative amount of the fully spliced transcript relative to the RI transcript increases, and the amount of the SI relative to the RI continues to increase throughout the rest of the time course.

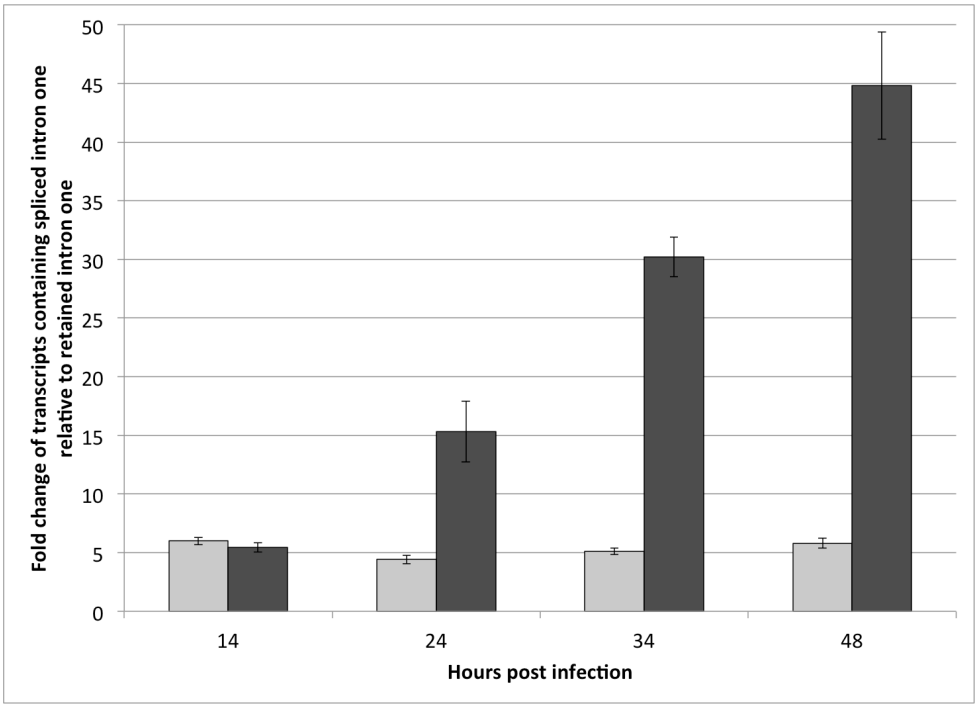


Figure 24. Intron one of *AT4G39270* is spliced out in response to infection with *B. cinerea*. RQ-RT PCR was used to quantitatively determine the proportion of the retained intron transcript compared to the spliced version over a 48-hour time course. The spliced version increases over the time course relative to the retained intron version in *B. cinerea* infected samples (dark grey bars), whereas in mock samples the proportions do not change (light grey bars). The amount of the spliced spliceform is significantly increased compared to mock samples at 24 hpi and later, at the 5% significance level (n=3, p-values 24 hpi = 0.01, 34 hpi p= 0.01, 48 hpi p=0.04), there is no significant difference in the relative proportions of transcripts in mock samples between the time points. Error bars indicate SEM of three biological replicates.

The Cooke (2013) data indicates that *AT4G39270* is also differentially expressed as well as being differentially alternatively spliced (DASed). To determine the overall amount of *AT4G39270* present, primers were designed in a region of the gene that was common to all isoforms. To account for variations in the amount of RNA between samples they were normalised to *PUX1* (*AT3G27310*) a gene known not to change in response to *B. cinerea* infection (Windram *et al*., 2012). The efficiency for these primers was 105% and 93% thus no Pfaffl correction was needed. The expression of the gene increases approximately five-fold at 24 hpi (compared to mock samples) and remains at this level at 34 hpi; there is no change in gene expression between mock and 14 hpi.

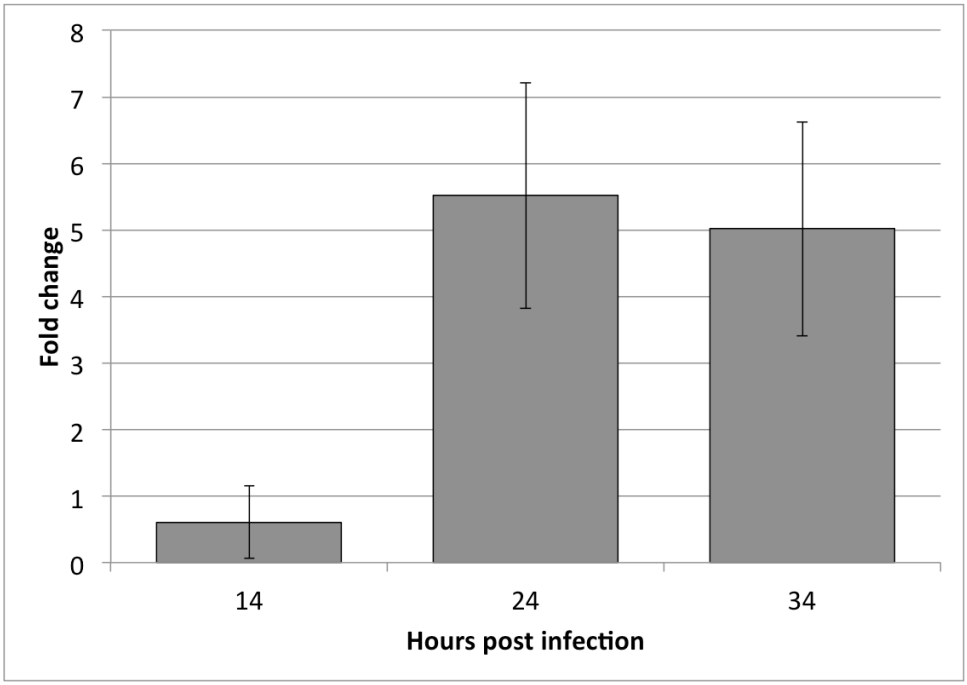


Figure 25. *AT4G39270* expression is increased in response to *B. cinerea* infection. At 14 hpi there is no increase in expression relative to mock samples, whereas at 24 hpi and 34 hpi there is approximately a five-fold increase in gene expression. Expression levels are normalised to *PUX* and relative fold increased compare to mock samples is shown, n=3

Taken together these results indicate that in response to *B. cinerea AT4G39270* expression is increased and the proportion of DASed transcripts alter, with the fully spliced transcript increasing relative to the RI transcript. Both the expression of the gene and the amount of the spliced spliceform increase up to 24 hpi. After that the expression level of *AT4G39270* appears to stay the same whereas the amount of the spliced transcript version relative to the retained intron version continues to increase. This indicates that the increase in the SI transcript relative to the RI transcript is not solely due to an increase in gene expression.

3.3.3.1 Differential alternative splicing of AT4G39270 is a plant driven defence response

Investigating if the classical fungal PAMP chitin is sufficient to produce DAS would help to confirm whether the DAS is a host or pathogen driven response. To determine if the presence of the pathogen is required for *B. cinerea*-mediated DAS of *AT4G39270*, four-week-old wild-type plants were spray-inoculated with Chitosan

(a chitin derivative) at three increasing concentrations (50mg/l, 100 mg/l and 150mg/l). At 2 hpi, leaves were detached and RNA was extracted. Using RQ-RT PCR, the proportion of transcripts where intron one was spliced relative to those where intron one was retained was determined using the primers detailed in Figure 23. In response to Chitosan, the amount of SI transcripts relative to the RI transcripts increased in a dose dependant manner (Figure 26). There is no statistical difference between mock and samples treated with 50mg/l of Chitosan. However, both the 100mg/l and 150mg/l treatments have an increase proportion of the SI transcript relative to the RI transcript, compared to mock samples (ANOVA $p<0.05$). The 150mg/l samples have a statistically significantly higher proportion of the SI relative to the RI, compared to all other samples (t-test $p<0.05$).

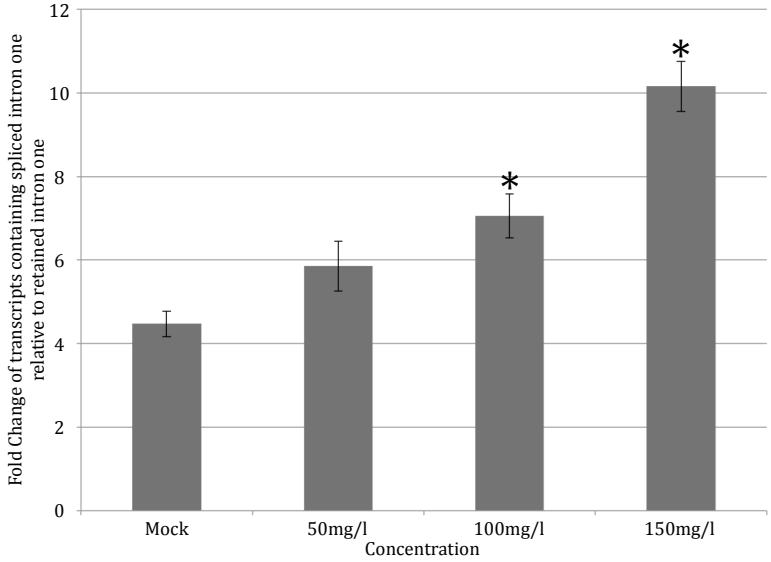


Figure 26. AT4G39270 is DAS in response to chitosan in a dose dependant manner. Application of 100mg/l and 150mg/l of chitosan results in a statistically significant increase in the spliced isoform relative to the retained intron isoform compared to mock samples, indicated by * (ANOVA $p<0.05$). Error bars indicate SEM of 4 biological replicates.

These results indicate the DAS of *AT4G39270* is a host driven defence response to *B. cinerea* infection. *AT4G39270* encodes a Leucine Rich Repeat Receptor Like Kinase (LRR-RLK); given the importance of LRR-RLKs in plant defence responses, we decided to further investigate this gene that undergoes *B. cinerea*-mediated DAS to determine if it displayed a defence phenotype.

3.3.4 The DAS gene, *AT4G39270*, is important in the plant defence response to *B. cinerea*

To determine if *AT4G39270* plays a role in the plant defence response to *B. cinerea*, three independent homozygous T-DNA KO lines of *AT4G39270* were obtained from the Nottingham Arabidopsis Stock Centre: SALK_072419C, SALK_051652C and SALK_047268C. The lines were screened for the presence of homozygous T-DNA insertion using gene-specific primers: SALK_072419C and SALK_051652C 3'-GAGGAAGACACAACGGTGTT-5' and 5'-AGCTCCAGAAGTATCTTGCC-3′ , and SALK_047268C 3'-CTGCAGTTGGAGGGTCTATATTG-5' and 5'-ACAGTAGCAAGCTTGTTTGTTAG-3', and the T-DNA specific primer SALK LBa1 5'-GCGTGGACCGCTTGCTGCAACT-3'. Each of these lines was confirmed to be homozygous for the T-DNA insertion (Figure 27).

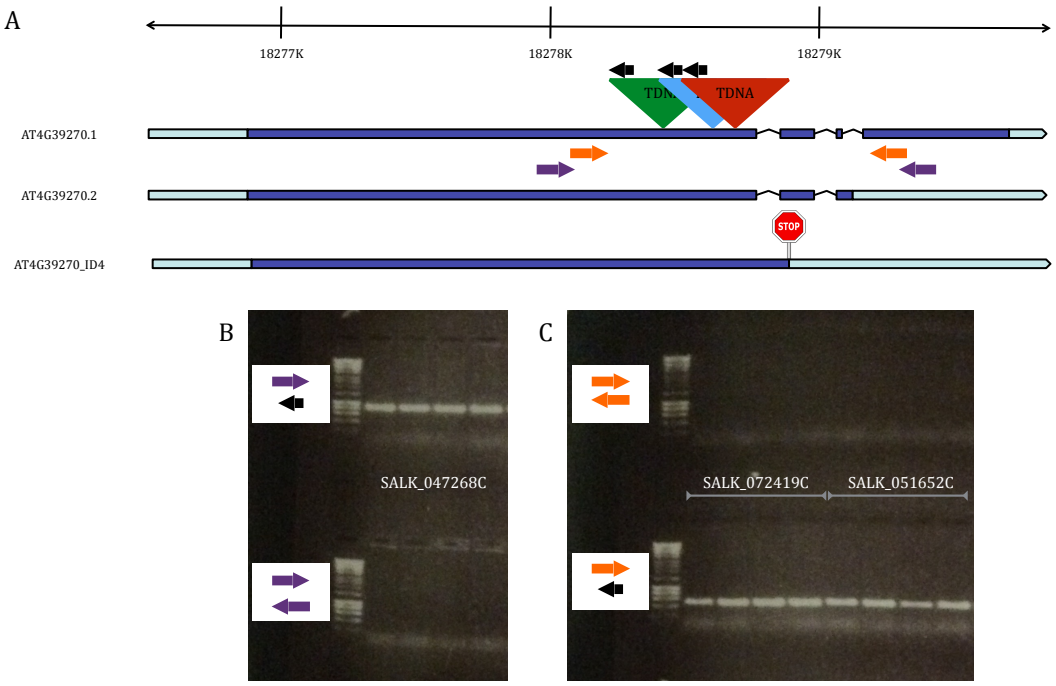


Figure 27. AT4G39270 KO lines are homozygous for T-DNA insertion. A. Location of T-DNA insertion and primers. Three independent T-DNA lines were tested, SALK_072419C, SALK_051652C and SALK_047268C represented by blue, red and green triangles. The insertion site in SALK_072419C and SALK_051652C are 47 bp apart thus the same gene specific primer could be used for both (represented by orange arrow); the purple arrows represent the gene-specific primers for SALK_047268C and the black arrows represent the T-DNA specific primer. B. SALK_047268C lines are homozygous. Representative PCR product gel of four samples. Top gene-specific forward primer and T-DNA specific reverse primer; bottom gene specific forward and reverse primer. C. SALK_072419C and SALK_051652C lines are homozygous. Representative PCR product gel of four samples (first four SALK_072419c, 5-8 SALK_051652C). Top gene-specific forward and reverse primer, bottom gene-specific forward primer and T-DNA specific reverse primer.

Of these three lines, two were morphologically similar to the *Arabidopsis* ecotype Columbia (Col0) wild-type, whereas the third line was smaller, flowered early and had lost its rosette appearance (Figure 28A). It also had significantly smaller leaves than the wild-type and either *AT4G39270-KO_1* (SALK_072419C) or *AT4G39270-KO_2* (SALK_051652C) (Figure 28B and C). It is possible that *AT4G39270-KO_3* (SALK_047268C) contains additional T-DNA insertions in another gene/s causing this dwarfed phenotype, due to the fact that T-DNA transformants contain an average of 1.4 inserts each (Feldmann, 1991). Thus only *AT4G39270-KO_1* and *AT4G39270-KO_2* lines were taken forward for further analysis.

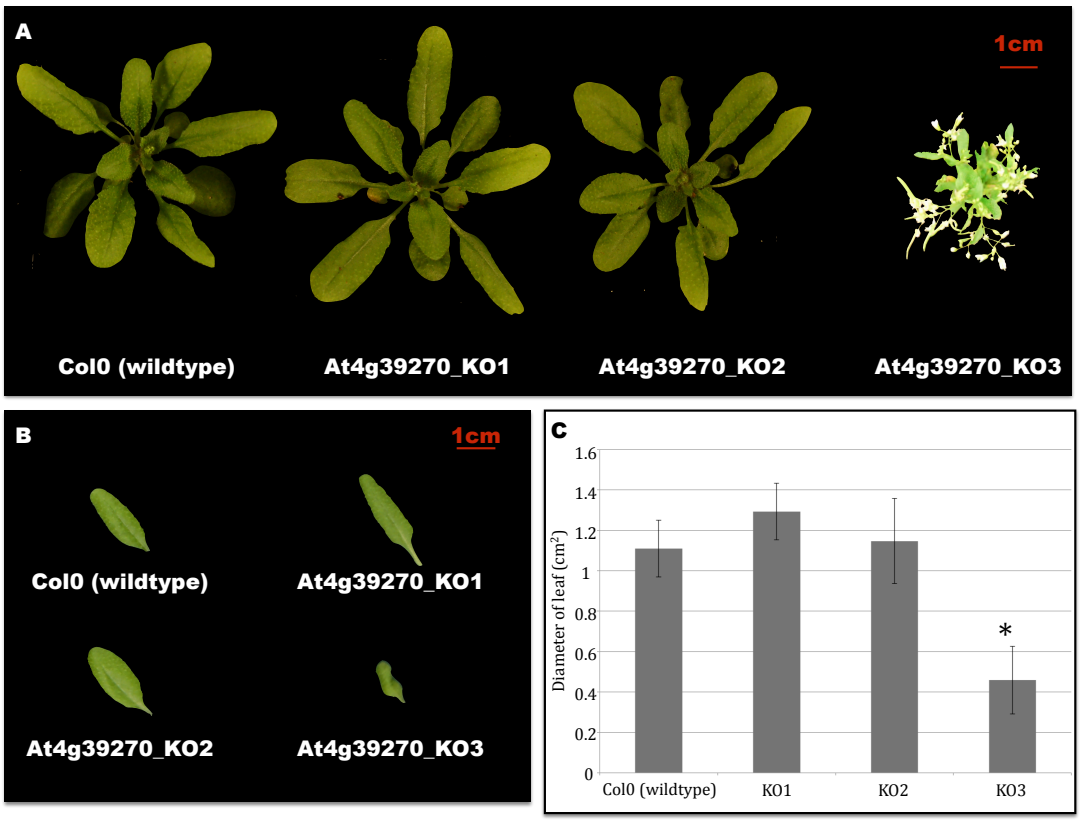


Figure 28. Morphology of AT4G39270 KO lines. Two independent KO lines of *AT4G39270* are similar in morphology to wild-type *Arabidopsis* plants. *A. AT4G39270-KO_1* (SALK_072419C) and *AT4G39270-KO_2* (SALK_051652C) have similar morphology to wild-type plants at five weeks. The third KO line (*AT4G39270-KO_3*) is smaller, flowers earlier and has lost the characteristic rosette of leaves. *B & C.* The leaf area of *AT4G39270-KO_1* and *AT4G39270-KO_2* are similar to wild-type, however *At4g93270_KO_3* (SALK_047268C) leaves are significantly smaller. Images and measurements taken using leaf 7 from four-week-old plants. *C* number of biological replicates=15, * indicates significantly smaller leaves compared to wildtype (Student T-test p<0.01), error bars SD.

To confirm that in these lines *AT4G39270* had been significantly knocked down / knocked out RQ-RT PCR was carried out on infected leaf samples; infected leaves were used because this increases the expression of *AT4G39270*. Both *AT4G39270-KO_1* and *AT4G39270-KO_2* were significantly knocked down, with both lines having less than 10% of the expression of wild-type plants (five biological replicates were used) (Figure 29).

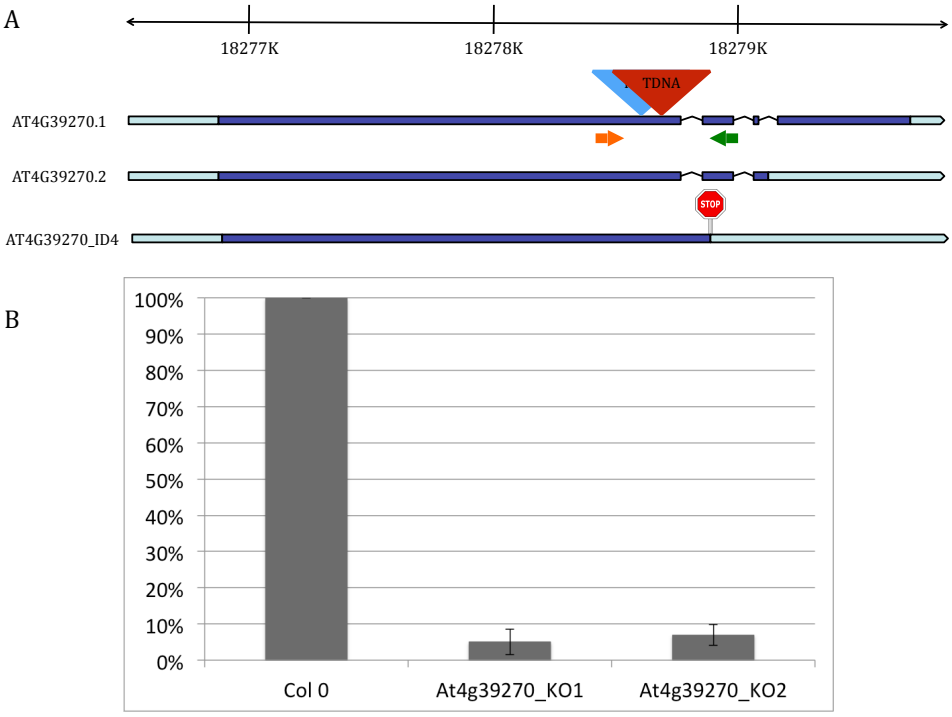


Figure 29. Gene expression of *AT4G39270* is knocked out in both T-DNA KO *AT4G39270-KO_1* and *AT4G39270-KO_2*. A. Location of T-DNA insertion for *AT4G39270_KO1* and *AT4G39270_KO2* represented by blue and red triangles respectively; orange and green arrows represent the location of the forward and reverse primers respectively. B. Gene expression of At4g392670 was determined using RT-RQ-RT PCR, expression relative to wild-type and normalised to *PUX*, n=5, error bars = SEM.

3.3.4.1 AT4G39270 loss-of-function mutants are more susceptible to B. cinerea infection.

The two T-DNA KO lines *AT4G39270-KO_1* and *AT4G39270-KO_2* plus a wild-type of the same background (Col0) were inoculated with spores of *B. cinerea* and the lesion size measured over a 72-hour time-period. Mock-inoculated leaves (leaves treated with only the inoculation medium) did not develop any lesions during this time-period. At 24 hpi there were no visible lesions. By 48 hpi lesions were visible in all lines and lesions continued to expand until the experiment ended at 72hpi. *AT4G39270-KO_1* and *AT4G39270-KO_2* developed significantly larger lesions than wild-type leaves (Figure 30). *Bos1,* a mutant known to have enhanced susceptibility to *B. cinerea* (Mengiste *et al*., 2003), was used as a positive control and as expected also developed significantly larger lesions than wild-type (Figure 30).

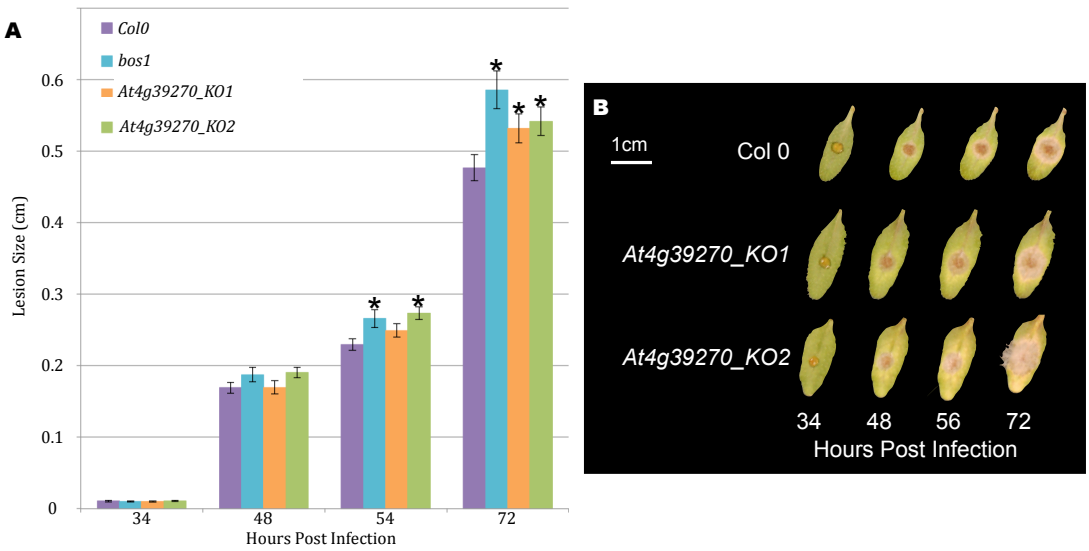


Figure 30. *AT4G39270-KO_1* and *AT4G39270-KO_2* are more susceptible to *B. cinerea* than wild-type plants. *A and B.* Leaf 8 of four week old *Arabidopsis* plants spot inoculated with *B. cinerea* show an increase in lesion size at times greater than 54 hpi for *bos1* and *Atg49370_KO2* lines and at 72hpi for *AT4G39270_KO1* compared to wild-type. * Indicates statistically significant (p<0.05 Students T-test), error bars indicate SEM of 30 biological replicates. B. The increasing lesion size on a representative leaf for wild-type (Col0), *AT4G39270-KO_1* and *AT4G39270-KO_2* lines over the course of infection is illustrated.

To confirm that the enhanced lesion development reflected increased growth of the pathogen, we measured the expression of the *B. cinerea* β-tubulin gene during infection of wild-type, *AT4G39270-KO_1* and *AT4G39270-KO_2* (Figure 31). This gene was used because it is a housekeeping gene, and as such can be used as a proxy for pathogen growth (Windram *et al*., 2012). Immediately after inoculation the levels of *B. cinerea* β-tubulin in leaves of all lines were not significantly different, as expected, given that leaves were inoculated with the same suspension of *B. cinerea* spores. The profile of *B. cinerea* β-tubulin expression in all lines shows the usual initial increase, lag phase and subsequent increase reflecting different stages of infection as seen by Windram *et al* (2012), with both *AT4G39270* KO lines and *bos1* showing statistically significant increase in fungal load compared to wild-type. The *B. cinerea* β-tubulin levels were not determined after 48 hpi because after this time point the fungus reaches the leaf boundary and can infiltrate the agar, thus potentially making results inaccurate.

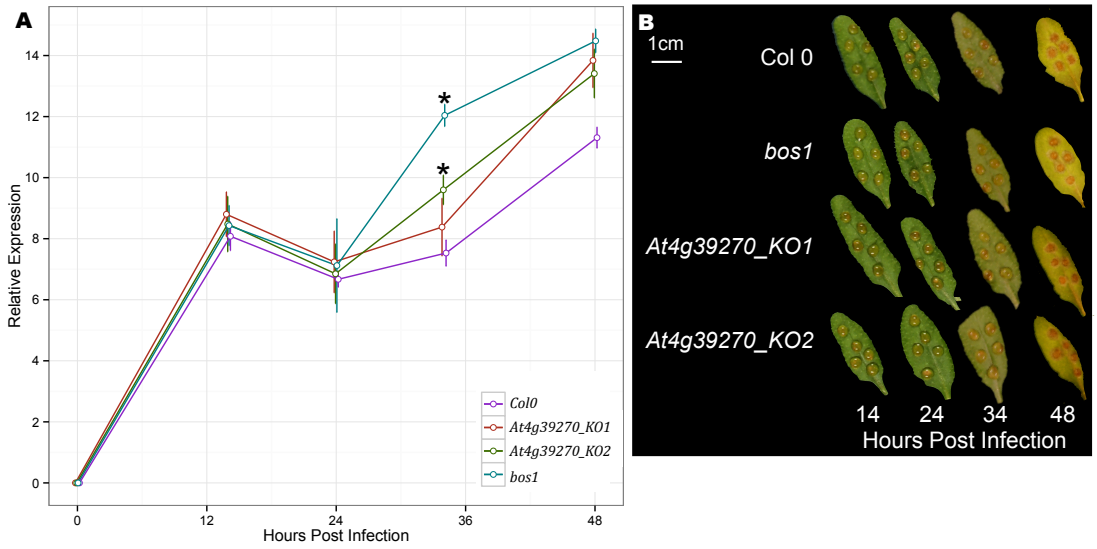


Figure 31. *AT4G39270* KO lines have increased fungal load compared to wild-type. A. *B. cinerea* β-tubulin gene expression levels were used as a measure of the growth of *B. cinerea* in leaf 7 of four-week-old *Arabidopsis* plants. This was determined using RT-RQ-RT PCR, shown as the log2 ratio of expression of *B. cinerea* β-tubulin relative to *Arabidopsis PUX1* gene (At3g27310). * indicates statistically significant increase in fungal load compared to wild-type samples at 34 hpi (Student t-test p<0.05), error bars indicate SEM of 6 biological replicates. *Bos1* and *Atg49370_KO2* lines display increased growth of *B. cinerea* at time points ≥ 34 hpi compared to wild-type, by 48 hpi *AT4G39270_KO1* also shows increased growth compared to wild-type. B. Detached leaves were inoculated with five 10µ droplets containing 10^5 spores per mL of *B. cinerea*. Visible lesions appear at 34 hpi. At 48 hpi the lesions have increased in size indicating that infection is progressing as expected.

These results indicate that the enhanced lesion development seen in *AT4G39270-KO_1* and *AT4G39270-KO_2* reflects an increased ability of the pathogen to grow in the host, suggesting that AT4G39270 plays an important role in the defence response against *B. cinerea*. Because both independent T-DNA KO lines of *AT4G39270* exhibit the same phenotype it strongly indicates that the *B. cinerea* phenotype displayed is specific to this gene being knocked out.

Taken together these results strongly indicate that *AT4G39270* plays an important role in Arabidopsis defence response against *B. cinerea*, with the DAS of *AT4G39270* being a plant driven response to infection. To determine if *AT4G39270* is important in a range of defence responses investigations into the susceptibility of the KO lines to biotrophic organisms was investigated.

3.3.4.2 AT4G39270 loss-of-function mutants are not more susceptible to biotrophic pathogens

To investigate if the LRR-RLK gene *AT4G39270* has a role in general pathogen response the susceptibility of *AT4G39270_KO1* and *AT4G39270_KO2* to two other pathogens, the obligate biotrophic oomycete *Hpa* (strain Nok1) and the hemi-biotrophic bacterial pathogen *Pst,* strain DC3000 was determined. Arabidopsis *AT4G39270_KO1* and *AT4G39270_KO2* lines display the same susceptibility to *Hpa* and *Pst, DC3000* as wild-type plants (Figure 32A and Figure 32B respectively). For both pathogens, transgenic Arabidopsis ecotype Col-0 were used as wild-type plants and a transgenic plant expressing an SA hydroxylase that degrades SA to catechol (*NahG*) was used as a positive control. This line has been shown to be more susceptible to *Hpa* and *Pst DC3000* than wild-type Col0 plants (Delaney *et al*., 1994). The *NahG* plants were significantly more susceptible to *Hpa* four days post infection (dpi) and plants spray-inoculated with *Pst DC3000* showed increased susceptibility from one dpi until the end of the time course, indicating that the assays worked and *AT4G39270* is not a vital gene for defence against *Hpa* or *Pst DC3000*.

This could indicate that *AT4G39270* either has a redundant function with other genes in response to infection with *Hpa* and *Pst*, or that it is not involved in the defence response to biotrophic organism, potentially implying that it could be a necrotrophic-specific defence-related gene.

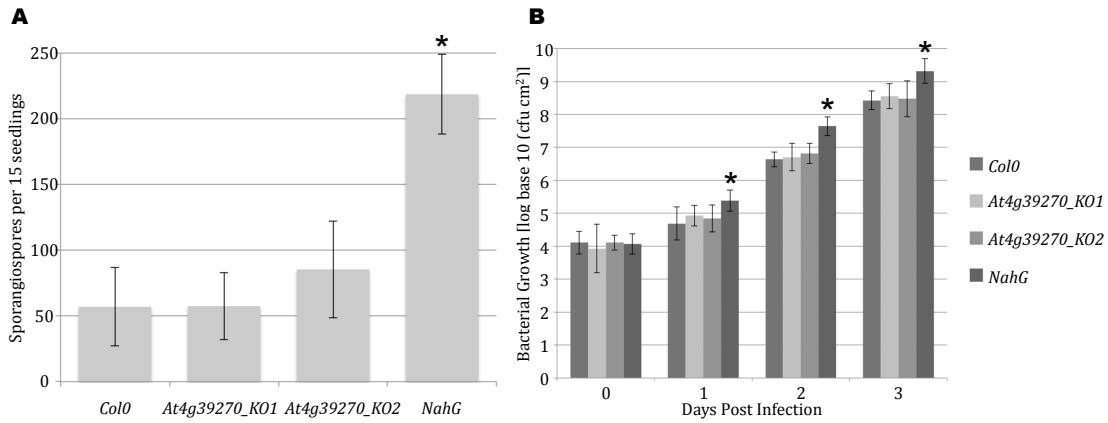


Figure 32. *AT4G39270* knocked out lines are not more susceptible to *HPA and Pst DC3000.* A. *HPA* phenotype screen. Two-week-old seedlings were spray-inoculated with a suspension of 30000 conidiospores per ml of *Hpa* isolate noks1. At four days post infection, the number of conidiophores per 15 seedlings was counted. Only the positive control, *NahG*, showed an increase in susceptibility compared to wild-type. Error bars indicate SD of 4 biological replicates, * indicate statistically significantly (Students T-test p < 0.01). B. *Pst* DC3000 phenotype screen. Four-week-old plants were spray-inoculated with virulent *Pst* DC3000, only the positive control, *NahG*, showed an increase in susceptibility compared to wild-type. Error bars indicate SD of 15 biological replicates, * indicate statistically significantly (Students T-test p < 0.01).

Taken together these results show that in response to *B. cinerea* infection intron one of *AT4G39270* undergoes DAS and that this gene is involved in the defence response to *B. cinerea* potentially in a pathogen specific manner.

3.3.5 Splicing regulatory elements are enriched in introns retained in response to *B. cinerea*

To investigate potential regulation mechanisms, the predominant type of event, RI, was investigated to determine if there are splicing regulatory elements (SRE) associated with the retention of introns in response to *B. cinerea* in Arabidopsis at the genome-wide level. The DNA architecture within RIs was analysed to determine if these areas were enriched for particular motifs. The nucleic acid sequence of each *B. cinerea*-mediated RI was extracted and the DNA architecture investigated to find putative SREs.

Mao *et al* (2014) divided all the introns in TAIR10 into a RI set, which consisted of introns that were retained in at least one isoform in the TAIR10 gene models, and constitutively spliced intron (CSI) set where the introns are spliced in all isoforms of the gene model in TAIR10. They found that in the RI set GGG-containing and GGAG-containing motifs are enriched, whereas AT/TA-rich motifs have a greater abundant in the CSIs set. This indicates that an increase in UA leads to more efficient splcing. Thus these motifs were chosen to be investigated further. The percentage of introns containing these motifs was determined for four datasets, *B. cinerea*-mediated RI (from the Cooke (2013) dataset), all introns in the TAIR10 genome, the TAIR10 RI and CSI classified TAIR10 introns by Mao *et al* 2014 (Cooke, 2013; Mao *et al*., 2014). In

addition two other purine rich motifs, GAAG, GAG were investigated because they have been implicated to act as regulators of AS that contribute to RIs in SR genes and heat mediated AS respectively (Chang *et al*., 2014; Thomas *et al*., 2012). Of these motifs, the most common motifs in *B. cinerea*-mediated RI were GAG and TAAT. In each case the percentage of introns was higher than in TAIR10 RI set indicating that the motif enrichment is not just a feature of an intron being able to be retained and thus potentially implying that they may be involved in *B. cinerea* mediated DAS. The least common motif in *B. cinerea*-mediated RI was GGAC, with the frequency being in line with that found in the other sets indicating that this motif may not be specific to *B. cinerea*-mediated DAS (Figure 33). Interestingly introns retained in response to *B. cinerea* contain a higher proportion of AT/TA-rich motifs than found in the all TAIR10 intron set, with the motifs TATAT and ATATT being found at roughly the same proportions as CSI and TAAT, TTATA and ATTAT being found in higher proportions, this could potentially indicate that AT/TA-rich motifs could be regulatory cis-elements in *B. cinerea*-regulated RI regions (Figure 33).

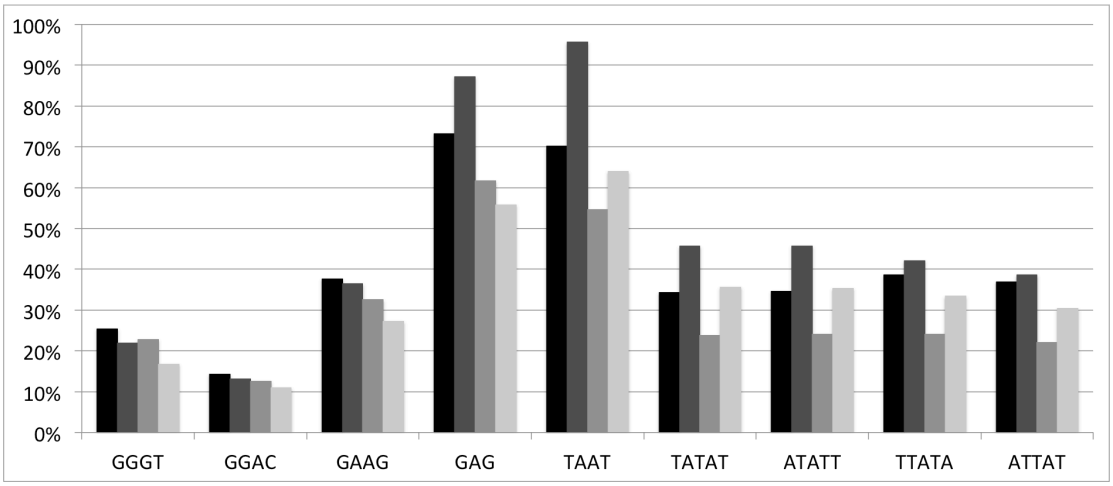


Figure 33. Introns retained in response to *B. cinerea* contain a higher proportion of GAG and AT/TA rich motifs compared to all RIs in TAIR10. GGGT, GGAG and GAAG motifs are found in approximately the same proportion in *B. cinerea* mediated RI, all TAIR10 introns and TAIR10 RI, whereas GAG and the AT/TAT rich motifs have a higher percentage abundance in *B. cinerea* mediated introns compared to RI in TAIR10. Percentage of introns containing motifs in *B. cinerea* mediated RI, all introns in TAIR10, TAIR10 RI set and constitutively spliced introns in TAIR10 are represented by black, dark grey, medium grey and light grey bars respectively.

To further investigate SRE motifs, DREME on CyVerse was utilised to determine which motifs were enriched compared to two different backgrounds, all TAIR10 introns and the TAIR10 RI set determined by Mao *et al* 2014 (Bailey, 2011; Goff *et al*., 2011; Mao *et al*., 2014). Of particular interest in elucidating *B. cinerea*-mediated DAS motifs associated with RI are those motifs that are enriched compared to both backgrounds, because these are motifs that appear in greater frequency than expected for Arabidopsis introns (the all introns in TAIR10 background) and potentially not just because according to TAIR10 gene models they can undergo an AS RI event (TAIR10 RI background). Thus they could help us to elucidate *B. cinerea*-mediated DASed RI associated motifs. Two different nucleotide length settings were used to detect motifs, hexamers and 9-mers, resulting in 29 and 32 hexamer motifs enriched compared to the TAIR10 intron and TAIR10 RI backgrounds respectively, with slightly fewer 9-mers detected (28 and 12) (Table 7). Two identical motifs were found to be enriched compared to both backgrounds in both the hexamers and in the 9-mer motif analysis (Table 7).

Table 7. Number of motifs found in introns retained in response to *B. cinerea* using the DREME software. Column one details the length of motif search for with columns 2-4 displaying the number of motifs found to be relatively enriched in the *B. cinerea*-mediated RI dataset compared to the background of all TAIR10 introns, TAIR10 RI and enriched in both backgrounds respectively.

	All TAIR10 introns	TAIR10 RI	Both sets
Hexamers	29	32	2
9mers	28	12	2
3-12mers	26	23	0

3.3.5.1 Hexamer motifs enriched in both backgrounds

The two hexamer motifs that were enriched in both backgrounds are STMTGA and CCTKAA (where S = A or T, M = A or C, K =G or T). To determine if these motifs have been identified as SRE previously, they were compared to the 84 computationally predicted ESE hexamers identified by Pertea *et al* (2007) in Arabidopsis, and the 238 ESEs identified by Fairbrother *et al* (2004) in humans. Neither of these two motifs matched ESE sequences in the Pertea *et al* (2007) dataset. However the CCTKAA

motif matches the ESE sequence CCTGAA (K represents either T or G) that is identified as a putative ESE in humans, and the consensus sequence of STMTGA from the TAIR10 RI background dataset (CTCTGA) matches a motif that has been shown to be a binding site for the human RBPs RNA-binding protein 6 (RBM6) and RNA-binding protein 10 (RBM10) (Bechara *et al*., 2013).

To further investigate motifs associated with *B. cinerea*-mediated DASed RI events, the consensus sequences of motifs that were enriched against both backgrounds were analysed. Six motifs have an enriched common consensus sequence compared to both backgrounds (Table 8).

Of these six, one, AAGCAA, has previously been predicted to be an ESE in Arabidopsis by Pertea *et al* (2007), and two others AGACAA and CAAAAG have been previously identified as ESEs in humans by Fairbrother *et al* (2004), potentially indicating that if they are present in an intron they could acts as ISEs.

Table 8. Six hexamer motifs that are enriched in *B. cinerea* retained introns compared to both TAIR10 RI and all TAIR10 introns share a common consensus sequence. Column one shows the consensus sequence and columns two and three show the *B. cinerea* motifs enriched compared to the TAIR10 RI set and all introns in TAIR10 respectively.

	Motifs	
Consensus sequence	TAIR10 RI	All introns in TAIR10
AAGCAA	WAGCAR 	AASCWA 
AGACAA	RGACAA 	AGACAM 

ATCTAA	AKCTAA 	ATCYAA 
ATTTGA	AYTKGA 	AYTTGA 
CAAAAG	CAHAAK 	CAVARG 
CCTTAA	CCTKAA 	CCTKAA 

3.3.5.2 Linking identified Splicing Regulatory Elements to RNA-Binding Proteins

Gosai *et al* (2015) carried out a global analysis of RNA-protein interaction in Arabidopsis seedling nuclei; using protein interaction profile sequencing (PIP-Seq). They identified protein-protected sites and utilized MEME and HOMER to identify 42 RBP-bound motifs. Comparing our motifs to these 42 could potentially indicate if our motifs are likely to bind RBPs. Comparing the consensus sequences that originated from motifs enriched in *B. cinerea*-mediated RI compared to both backgrounds with this dataset, identified two motifs that are also RBP-bound motifs AGACAA and AYTKGA. AGACAA, as mentioned above, is predicted to be an ESE by Pertea *et al* (2007) and forms part of the TGCTGGAGACAA RBP-bound motif detected by Gosai *et al* (2015). The consensus sequence, ATTTGA, originating from the motifs AYTKGA and AYTTGA, is found within the RBP-bound motif GGAAUUUGAGAA (U replaces T in RBP-bound motifs because they were working with RNA sequences). This ATTTGA motif was also part of a 9mer that was enriched in *B. cinerea*-mediated RI compared

to the background of TAIR10 RI, which also has a similar sequence to the RBP-bound motif (9-mer motif AA[C/T]TTGA[A/C]A) (Figure 34).

B. cinerea-mediated RI hexamer motif

B. cinerea-mediated RI 9mer motif

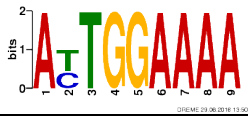
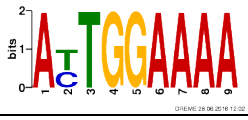
RBP-bound motif

Figure 34. The motif ATTTGA is a putative cis regulatory element involved in *B. cinerea*-mediated RI events. Top Panel: the ATTTGA motif was identified as relatively enriched in *B. cinerea*-mediated RI using the DREME software via CyVerse (Bailey, 2011; Goff *et al*., 2011). Middle panel: it was also identified as part of a 9mer that we determined to be relatively enriched in *B. cinerea*-mediated RI compared to the background of TAIR10 RIs. Bottom panel: both of these motifs are found within a RBP-bound motif identified by Gosai *et al* (2015). The hexamer consensus sequence is an exact match (black line), whereas the 9mer motif matches all positions however one is a poor match with A replacing G (red line) (blue line and green lines represent 9mer consensus sequence match, and good motif match respectively).

3.3.5.3 9mer motifs enriched in both backgrounds

Two 9mer motifs were enriched in *B. cinerea*-mediated DASed RI events compared to both backgrounds, AYTGGAAAA and ATGCATCAD. In addition one consensus sequences, AAAAATCAA, was enriched compared to both backgrounds. However, it originates from different motifs depending upon the background. In the TAIR10 intron background the enriched motif is ADAAATCA, whereas in the TAIR10 RI background it is AWAAATCAA (Table 9).

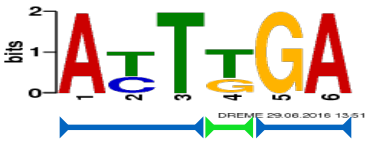
Table 9. Three 9mer motifs that are enriched in *B. cinerea* retained introns compared to both TAIR10 RI and all TAIR10 introns that share a common consensus sequence. Column one shows the consensus sequence and columns two and three show the *B. cinerea* motifs enriched compared to the TAIR10 RI set and all introns in TAIR10 respectively.

	Motifs	
Consensus sequence	TAIR10 RI	All introns in TAIR10
ATTGGAAAA	AYTGGAAAA 	AYTGGAAAA 
ATGCATCAT	ATGCATCAD 	ATGCATCAD 
AAAAATCAA	AWAAATCAA 	ADAAATCAA 

The first six positions of the *B. cinerea*-mediated RI motif AYTGGAAAA match the *B. cinerea*-mediated RI hexamer AYTKGA, in addition two RBP-bound motifs share sequence homology with this motif (Figure 35).

A *B. cinerea*-mediated RI 9mer motif

B *B. cinerea*-mediated RI hexameter motif



C RBP-bound motif 1

D RBP-bound motif 2

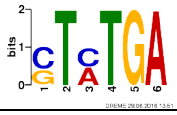
Figure 35. The motif AYTGGAAAA is a putative cis regulatory element involved in *B. cinerea*-mediated RI events. A. The AYTGGAAAA motif was identified as enriched in *B. cinerea*-mediated RI using the DREME software via CyVerse (Bailey, 2011; Goff *et al*., 2011). B. Part of it also shares sequence similarities with the AYTKGA that we determined to be enriched in *B. cinerea*-mediated RI relative to the background of TAIR10 RIs. C-D. The AYTGGAAAA motif shares sequence similarities with two RBP-bound motifs identified by Gosai *et al* (2015), blue line and green lines represent 9mer consensus sequence match, and good motif match respectively, with the red line indicating no match.

Gosai *et al* (2015) identified two genes that interact with the RBP-bound motif UAAAGGUGGAAA at a greater than six-fold increase, AT4G22960 (DUF544) and AT1G76010 (AlbaRBP), potentially indicating a role for these in *B. cinerea* mediated DAS.

3.3.5.4 Putative cis-elements are found in validated RI events

Out of the five genes with validated RI events three of them contain motifs detected as enriched in *B. cinerea*-mediated RI Table 10**Error! Reference source not found.**.

Table 10. *B. cinerea*-mediated RI motif sequences are present in the introns of some validated *B. cinerea*-mediated DASed genes. Column one identifies the motif, and columns 2-5 details the sequences found in the validated genes (column 2-5).

	Genes			
Motif	AT1G74590	AT4G39270 intron 1	AT4G39270 intron 2	AT5G16800
WAGCAR 	TAGCAG			
AKCTAA 			ATCTAA	AGCTAA
STMTGA 		CTATGA	GTATGA	

Interestingly both intron one and intron 2 contain a STMTGA motif sequence but in intron one the sequence is CTATGA and in intron two it is GTATGA, this could potentially result in the same RBP binding to each region but with slightly different strength. *AT4G39270* contains more than one putative SRE, this is not unexpected because in mammals multiple SREs, along with their corresponding splicing factors, act together to determine the finial splicing outcome of a gene (Blencowe, 2006).

3.4 Discussion

AS enables organisms to increase the diversity of their proteasome and can help regulate gene expression through a variety of mechanisms including but not limited to, affecting transcript stability when it is coupled to NMD (Reddy *et al*., 2013). A variety of studies have shown that DAS of individual or families of genes plays an important role in regulating the plant defence response. For example DAS of R genes (Dangl & Jones, 2001; Dinesh-Kumar & Baker, 2000; Jens Staal, 2008; Weaver *et al*., 2006; X.-C. Zhang & Gassmann, 2003). However few studies investigating DAS at the genome-wide level have been undertaken. Previously most genome-wide studies were performed on steady state mRNA rather than investigating isoforms, with specially designed tilling arrays required to investigate the latter. With the development of RNA-Seq, whole genome AS studies can now be performed relatively easily and cheaply helping to elucidate how DAS is involved in the defence response. Work carried out by Cooke (2013) strongly indicates that genome-wide DAS occurs in Arabidopsis leaves in response to infection with *B. cinerea* and that this may play a functional role in plant defence (Cooke, 2013). However, the Cooke (2013) dataset investigated one biological replicate per time point, and because AS can be a consequence of the stochastic nature of the splicing machinery more biological replicates are required in order to be able to draw conclusions at the plant population level.

3.4.1 The Cooke RNA-Seq dataset is robust

Although the lack of biological replicates is a major drawback of the Cooke (2013) dataset it compared favourably to the Windram *et al* 2012 dataset with respect to detecting DEG and had a reasonable validation rate for *B. cinerea*-mediated DAS genes, indicating that plant gene responses to *B. cinerea* can be detected by RNA-Seq with minimal replicates (Cooke, 2013; Windram *et al*., 2012). This gives us confidence that those genes Cooke (2013) detected to be DAS actually are. The RNA-Seq data covered a large dynamic range of differential gene expression and is potentially more sensitive at detecting DEG then microarray experiments. Detection of DAS transcript changes requires more sensitivity than detecting DEG because they

are of lower abundance, showing that the Cooke dataset detects lowly expressed DEGs indicates that it has the potential to detect DAS transcripts.

3.4.1.1 The Cooke differentially alternatively spliced dataset is consistent with published works

Cooke (2013) identified 694 *B. cinerea*-mediated DAS events, which can be divided into six categories, with the proportion of these event types being in line with previously published work. This indicates that all types of events are identified in this dataset without bias towards a particular event type. The proportion of RI events is slightly higher than published works on AS but in line with work published of DAS (Ding *et al*., 2014; Howard *et al*., 2013; Iida *et al*., 2009; Marquez *et al*., 2012; McGuire *et al*., 2008; B.-B. Wang & Brendel, 2004a).

3.4.1.2 B. cinerea-mediated retained intron events may be a defence regulatory mechanism

To determine if the increased proportion of RI events seen in the Cooke (2013) dataset compared to AS datasets could potentially have a biological function in the stress response genes containing *B. cinerea*-mediated RI events were investigated further. Genes containing *B. cinerea*-mediated RI events identified by Cooke (2013) showed that for those genes containing at least one additional intron, transcripts where at least one of these intron was spliced was detected in 96% of cases. This strongly indicates that the RI events seen are not due to detecting partially processed transcripts and could potentially imply that retention of introns is a plant defence mechanism. This is consistent with the hypothesis presented by Wachter (2014) where in uninfected tissues R genes are turned over by AS. Upon infection NMD is supressed resulting in an increased expression of R gene isoforms, which leads to resistance. Because the Cooke (2013) dataset comprises of only one biological replicate the RI events could be due to noise generated by the splicing machinery, RNA-Seq experiments with a higher number of biological replicates could help to determine if increase in the number of RI events in response to *B. cinerea* infection could be a defence response mechanism. Evidence that supports the idea

that an increase in the number of RI events is involved in the plant defence mechanism comes from published work where in both salt stress and *Pst* infection experiments where there are a higher number of RI events detected (70 and 67% respectively) than in AS studies carried out under non stress conditions (39%-56% of events) (Ding *et al*., 2014; Howard *et al*., 2013; Iida *et al*., 2009; Marquez *et al*., 2012; McGuire *et al*., 2008; B.-B. Wang & Brendel, 2004a).

3.4.1.3 Differential Alternative Splicing can occur in both a specific and general stress responsive manner

Relatively few genome-wide DAS studies investigating the stress response have been undertaken. Comparing *B. cinerea*-mediated DASed genes with those determined to be DASed in two such studies investigating an abiotic and biotic stress (salt stress and *Pst* infection respectively) indicates that there are some genes that are DASed in a stress specific manner and some that undergo DAS as a more general stress response. This is in line with what is seen in gene expression data. Sham *et al* (2015) compared publically available *B. cinerea* microarray data with heat, salt or osmotic stress and determined that some stress responses are conserved between all stresses, potentially conferring tolerance to many different stresses, and some are stress specific (Sham *et al*., 2015). One drawback to all three of these RNA-Seq studies was the limited number of biological replicates used.

Although the *B. cinerea*-mediated DAS dataset by Cooke (2013) has only one biological replicate it appears to identify biologically relevant DASed genes. The validation rate of a subset of these genes in new biological replicates is 76.5%. This combined with functional analysis results that showed *B. cinerea*-mediated genes are enriched for defence related GO terms, indicates that it is a relatively robust dataset that could be utilised to investigate *B. cinerea*-mediated DAS.

3.4.2 Differentially Alternatively Spliced marker genes are identified

Seven genes were validated to undergo *B. cinerea*-mediated DAS. The seven genes undergo a range of AS event types and cover a diverse range of gene families

(Cooke, 2013). In addition, some of these genes have been detected to undergo DAS in response to more than one stress; *AT4G39270* undergoes DAS in response to *B. cinerea* and *Pst* infection (with different DAS events) whilst *AT1G28330*, *AT1G74590* and *AT3G02300* are DASed in response to both *B. cinerea* infection and salt stress (with the same event), with the remaining four only undergoing DAS in response to *B. cinerea* potentially indicating that they undergo stress specific DAS. Thus these genes represent a diverse set of DASed events that could be used as marker genes in further DAS studies.

3.4.3 The *B. cinerea*-mediated differentially alternatively spliced gene, *AT4G39270* has a defence phenotype

We have demonstrated that the LRR-RLK encoding gene, *AT4G39270*, is an important gene for the defence response to *B. cinerea* and that it undergoes *B. cinerea*-mediated DAS. LRR-RLKs have been shown to be involved in the perception of pathogen attack, and in transducing this signal downstream to the mitogen activated protein kinases. We hypothesise that AT4G39270 is involved in the detection of *B. cinerea*, by either directly perceiving a MAMP or PAMP produced by the pathogen and thus acting as a pattern recognition receptor (PRR,) or by associating with a PRR to transduce the signal to the MAP kinases. Further work would be required to confirm its role in the defence response.

3.4.3.1 The first intron of AT4G39270 is retained in response to B. cinerea infection

The Cooke (2013) RNA-Seq dataset predicts that upon infection with *B. cinerea* the isoforms where intron one is spliced (AT4G39270.1 and AT4G39270.2) increases in proportion compared to *AT4G39270_ID4*, which has intron one retained. On average 98% of transcripts that encode functional protein are AT439270.1 with the other 2% being AT4G39270.2. Interestingly Howard *et al* (2013) have shown that the AS event between *AT4G39270.1* and *AT4G39270.2* is differentially spliced in response to *Pst* potentially indicating that different RI events occur in a pathogen specific manner. Both of the spliceforms have an active kinase domain with *AT4G39270.2* containing

an extra LRR domain due to the retention of the third intron. The *AT4G39270_ID4* spliceform has all the introns retained, with the retention of the first intron resulting in a premature termination codon (PTC). DAS of this first intron was confirmed using RT RQ-RT PCR, showing that in response to *B. cinerea* infection the proportion of transcripts where intron one is spliced increases relative to those where intron one is retained. This indicates that *AT4G39270* undergoes *B. cinerea*-mediated DAS where the proportion of the spliced transcripts (*AT4G39270.1* and *AT4G39270.2*) increases relative to *AT4G39270_ID4* with the proportion of *AT4G39270.1* compared to *AT4G39270.2* remaining unchanged.

The DAS of *AT4G39270_ID4* could be a mechanism that helps to regulate gene expression, potentially by being coupled to NMD. Kalyna *et al* (2012) have shown that many genes important for plant development and adaptation including transcription factors, RNA processing factors and stress response genes undergo regulation via AS coupled to NMD. In uninfected samples the *AT4G39270_ID4* transcript could be targeted for NMD due to the PTC creating a long 3'UTR. Upon infection with *B. cinerea* DAS occurs and the proportion of *AT4G39270_ID4* compared to the transcripts encoding the functional protein isoforms decreases in a time dependant manner. This could result in fewer transcripts undergoing NMD thus stabilising and increase gene expression. However Kalyna *et al* (2012) determined that the majority of PTC resulting from RI events that possess NMD triggers are insensitive to NMD, with Gohring *et al* (2014) indicating that this is because they are retained in the nucleus. Thus, it is more likely that the transcripts of *AT4G39270_ID4* are retained in the nucleus and thus are not targeted by NMD, however they would also be unable to be translated into proteins (because translation occurs outside the nucleus) thus DAS of *AT4G39270* could still potentially influence the amount of functional protein produced. One way to determine if *AT4G39270_ID4* is targeted for NMD would be to determine if the ratio of spliced intron one transcripts relative to retained intron one transcripts is altered in NMD mutants compared to wild-type plants.

The function of *AT4G39270* is unknown. However, due to its protein structure (LRR N-terminal and a kinase domain C-terminal) it could potentially encode a PRR whose expression is in part regulated by *B. cinerea*-mediated DAS (potentially by AS coupled to NMD or another mechanism). In response to infection with *B. cinerea* more of the functional transcript is produced, which results in more protein, thus making the plant more sensitive to detecting *B. cinerea*. Although this is speculative there is evidence that PAMPs can trigger enhanced expression of PRR with FLS2, ERF and WAK1 all being up-regulated in response to perception of their own ligands (Brutus *et al*., 2010).

3.4.3.2 Loss-of-function mutants of AT4G39270 have increased susceptibility to B. cinerea

B. cinerea phenotype screens demonstrated that AT4G39270 KO lines have an increased susceptibility to the pathogen, with RT-PCR analysis of the amount of pathogen β-tubulin indicating that this is due to an increase in fungal growth. Ideally, to determine if DAS of this gene is important and whether transcripts with the first intron retained and spliced are both required for an effective plant defence response, transgenic plants that only produce one transcript should be made and the susceptibility of *B. cinerea* investigated. However, the transcript where intron one is retained has the same sequence as genomic DNA, thus it cannot be silenced on its own because the transcripts that are predicted to encode the functional transcripts where intron one is spliced can be made from this transcript. We considered mutating the splice sites to prevent splicing from occurring. However this can lead to cryptic splice sites. Utilising the web-based tool SlicePort both cryptic donor and acceptor site were predicted to occur in intron one (Figure 36) (Dogan *et al*., 2007). Thus if the splice site were mutated abnormal splicing is likely to occur.

CACAGgtatgaaaaaccattccttttactctttttttactttcaagccaccattagttcattactgatcttgagtcttccaatgttcaatagGGATA

Figure 36. Cryptic splice sites are predicted to occur in intron one. SlicePort, a web-based tool for detecting splice sites predicts that intron one of AT4G39270 contains cryptic donor (underlined in yellow) and acceptor (green line above) sites. Yellow and green box highlight the standard donor and acceptor sites respectively (Dogan *et al*., 2007).

3.4.3.3 AT4G39270 in Defence against biotrophic pathogens.

Interestingly *AT4G39270* KO lines did not affect the ability of Arabidopsis plants to mount an effective defence response against either the obligate biotrophic oomycete pathogen *Hpa* strain Nok1 or the hemi-biotrophic bacteria pathogen *Pst* strain DC3000. Although *AT4G39270* KO lines did not affect the defence response against *Pst* it has been shown to be induced in response to flg22 (Navarro *et al*., 2004). However Howard *et al* (2013) found that *AT4G39270* is not differentially expressed in response to either the virulent *Pst* DC3000 or the avirulent strain *Pst* DC3000 expressing the bacterial effector AvrRps4, but they observed a 37% increase in *AT4G39270.2* compared to *AT4G39270.1* spliceforms in plants inoculated with the avirulent strain compared to the virulent strain at 1 hpi. This could indicate that *Pst*-mediated DAS of *AT4G39720* results in intron three being retained compared to *B. cinerea*-mediated DAS where it is intron one that undergoes DAS. This potentially indicates that DAS can result in the same gene producing different proportions of transcripts in a stress dependant manner. If our speculative hypothesis that *AT4G39270* encodes a PAMP receptor is correct an increase in the *AT4G39270.2* transcript could enable better recognition of the virulent strain because the retention of intron three effects the number of LRR repeats. Why knocking out this gene has no affect upon susceptibility to *Pst* is unclear. However it may be due to redundancy of other genes. Multiple PRRs are activated by pathogens thus if AT4G39270 is acting as a PRR one or more of these could potentially compensate for the loss of *AT4G39270*. In addition, it could depend upon what PAMP it is detecting, for example if it is detecting a fungal specific PAMP then loss of AT4G39270 may not impact upon infection by bacteria. Thus these results indicate that different pathogens may affect DAS differently, potentially due to altering splicing factor concentrations.

3.4.4 Ten Receptor Like Kinase genes are Differentially Alternatively Spliced in response to *B. cinerea*

AT4G39270 is part of a large family of receptor like kinases with 610 members in *Arabidopsis*, with the LRR-RLK being one of the largest subfamilies (Shiu & Bleecker,

2001; Shiu *et al*., 2004). A subset of these LRR-RLKs containing 33 genes were analysed with the majority showing AS transcripts in RNA extracted from whole Arabidopsis plants (Gou *et al*., 2010). Mining the Cooke (2013) RNA-Seq dataset identifies ten RLK encoding genes that are DAS (from 12 different splicing events) in response to *B. cinerea* infection (Table 11).

Table 11. Genes from the LRR-RLK family that undergo differential alternative splicing in response to *B. cinerea* infection.

Gene name	Subfamily	Novel AS event	AS event in TAIR10
At1g18390	LRK10L-1	AFE	AFE
At1g21270	WAKL	RI	
At1g66880	LRK10L-1	AFE	
At1g73080	LRR XI	RI	
At2g23450	WAKL	RI	
AT3g21630	LysM	RI	
AT3g49060	RLCK IX	RI	
AT3g59350	RLCKVIII		AFE
AT4g23150	DUF26	A3SS	
AT4G39270	N.A.	RI	RI

Nine of these events represent novel transcripts of these genes, with the remaining 3 isoforms already annotated in TAIR10 genome release. As in the case of *AT4G39270*, all 10 of these RLK genes are also differentially expressed in response to *B. cinerea*. Three of these 10 RLKs are known to play a role in plant-pathogen interactions, At1g21270 (WAK2), At1g73080 (ATPEPR1) and AT3g21630 (ATCERK1) (Kohorn *et al*., 2012; Krol *et al*., 2010; Miya *et al*., 2007). Of particular note is ATCERK1 a known chitin PRR. Zhang *et al* (2014) identified two splicing factors, SUA (SUPPRESSOR OF ABI3-5) and RSN2 (REQUIRED FOR SNC4-1D 2), that are required for the proper splicing of *CERK1* and the atypical RLK *SNC4*.

Of the ten RLK genes DAS in response to *B. cinerea* four genes undergo *Pst*-mediated DAS, *AT1G66880*, *AT3G39350*, *AT4G23150*, and *AT4G39270*. One of which has been shown to be involved in defence related processes; *AT4G23150* encodes cysteine-rich receptor like kinase 7 (CRK7), which has been shown to mediate the response to extracellular reactive oxygen species (ROS) production in Arabidopsis (Nheimo *et al*., 2014) These results taken together indicate that DAS of RLKs is an important defence response and is potentially involved in PTI in response to *B. cinerea* infection.

3.4.5 Putative regulatory motifs are detected in *B. cinerea*-mediated retained introns

Investigating the genomic sequences of introns retained in response to *B. cinerea* in the Cooke (2013) dataset identified a number of putative cis-elements that could be involved in DAS of introns in response to stress.

3.4.5.1 GAG rich motifs are putative cis-elements related to stress mediated differential alternative splicing

Utilising the Cooke (2013) dataset we determined that introns that are retained in response to *B. cinerea* had a higher proportion of GAG rich areas than either all introns in TAIR10, or introns in TAIR10 that in at least one gene model they are retained. This could potentially indicate that GAG motifs play a role in the *B. cinerea*-mediated responses. GAG repetitive motifs have previously been identified by Chang *et al* (2004) to be regulators of AS that contribute to RI in heat mediated DAS, potentially indicating that this motif modulates RI in a generalised stress response.

To further investigate motifs in the *B. cinerea* RIs, DREME was utilised to identify motifs enrichment compared to all TAIR10 introns and TAIR10 introns where in at least one gene model the intron is retained (referred to from now on as the TAIR10 RI set) (Bailey, 2011). No motifs enriched for GAG repeats were discovered, but these motifs were determined from *B. cinerea* mediated DAS RI events detected using only one biological replicate. This could make the data noisy due to some AS

RIs being stochastic in nature. Thus investigating these motifs in a *B. cinerea* mediated DAS RI dataset originating from more biological replicates that are analysed to reduce false positives would be advantageous in identifying more motifs. Because the false positives should be random in nature, it should not detract from the accuracy of motifs detected, it just means that we might not detect some motifs due to the decreased power of the analysis.

3.4.5.2 Known Exon Splicing Elements cis-elements are detected in B. cinerea-mediated Retained introns

A previously computationally predicted ESE, AAGCAA, was identified as the consensus sequence of two *B. cinerea*-mediated RI enriched motifs (Pertea *et al*., 2007); motif WAGCAR was enriched compared to all TAIR10 introns and AASCWA compared to TAIR10 RI set. Only two hexamer motifs were identified to be enriched compared to both backgrounds, STMTGA and CCTKAA, one of which, CCTKAA, consensus sequence has been identified as a putative ESE in humans. Two other consensus sequences originating from motifs enriched in *B. cinerea*-mediated RIs but with different motifs depending on which background was used, AGACAA and CAAAAG, are also predicted to be ESEs in humans by Fairbrother *et al* 2004. These results indicate that ESE motifs when located in introns can act as ISEs. This has been indicated to be the case in humans where work by Wang *et al* 2012 indicates that the overlap between ISS and ESE activity could be a general phenomenon. Further evidence of this comes from the *B. cinerea*-mediated RI motifs (enriched compared to TAIR10 introns), AGMAGAAGA and GAAGAAKCA, both of which contain GAAG motifs. The GAAG motif is also found as part of the experimentally verified 9mer, GAAGAAGAA, found to be an ESE in plants (Pertea *et al*., 2007). In addition the GAAG motif has also been implicated as a SRE in humans with a repetitive GAAG motif being identified as a putative intronic splicing regulatory motif of the human dystrophin gene that results in an alternative exon event (Gualandi *et al*., 2003). The GAAG-containing motifs were not found in the TAIR10 RI dataset, potentially indicating that this motif enables AS of introns but is not regulated in a *B. cinerea*-mediated manner.

3.4.5.3 Putative B. cinerea-mediated retained intron motifs are linked to RNA-Binding Proteins

Two hexamer consensus sequences from *B. cinerea*-mediated RI motifs shared sequence similarities with RBP-bound motifs identified by Gosai *et al* (2015), AGACAA and ATTTGA. One RBP motif, TGCTGGAGACAA, contains within it the AGACAA motif that is enriched in *B. cinerea*-mediated RI sequences. The motif AGACAA also forms part of the human SRp38 consensus recognition sequence (AAAGACAAA) (Shin & Manley, 2002). Although the SR protein SRp38 is not present in plants, Richardson *et al* 2011 performed a comparative analysis of SR proteins across 27 eukaryotes and determined that SRp38 sub-family is a close sister group to the plant specific SCL subfamily.

Out of the four SCL members in Arabidopsis two, *AtSCL30* (*AT3G55460*) and *AtSCL33* (*AT1G55310*), are differentially expressed in response to *B. cinerea* at multiple time points in the Windram *et al* (2012) dataset, with *AtSCL30* being up-regulated and *AtSCL33* being down-regulated Figure 37. This could potentially indicate that DE of these genes could help regulate *B. cinerea* mediated DAS with the AGACAA motif binding one or more members of this family. However, further work including binding experiments would need to be carried out to determine if this is the case.

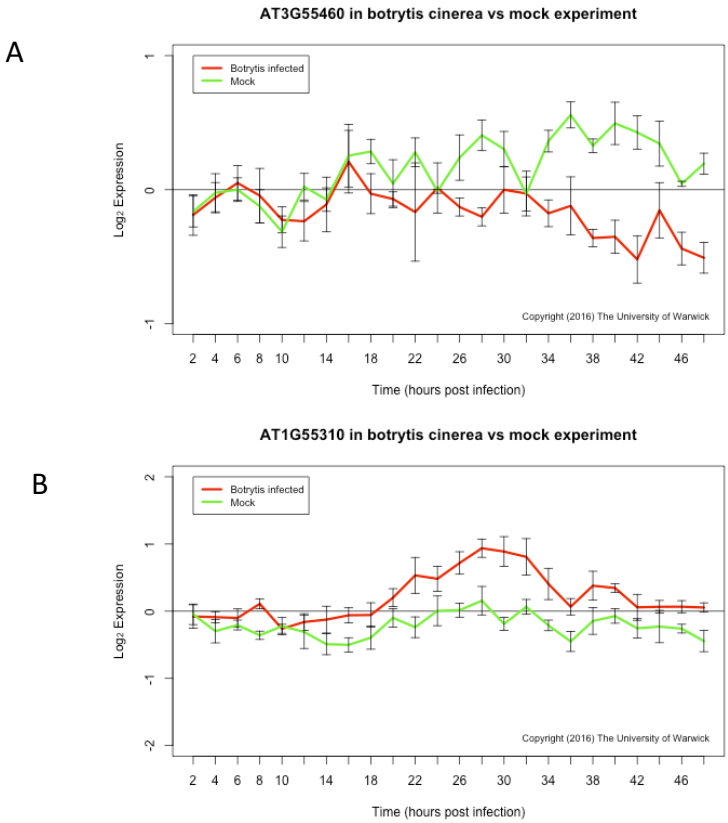


Figure 37. AtSCL30 and AtSCL33 are differentially expressed in response to *B. cinerea* infection. A. AtSCL30 (AT3G55460) is down-regulated in response to infection, whereas, B. AtSCL33 (AT1G55310) is up-regulated. Red line indicates *B. cinerea* infected and green line represents mock samples, images obtained via the gene expression profile viewer of the Windram *et al* (2012) data.

One 9-mer, AYTGGAAAA, shared sequence similarities with two RBP-bound motifs, one of which, UAAAGGUGGAAA, Gosai *et al* (2015) identified as interacting with two genes *AT4G22960* (*DUF544*) and *AT1G76010* (*AlbaRBP*). Windram *et al* (2012) identified *AT4G22960* as being DE in response to *B. cinerea* potentially indicating that this gene could be involved in mediating DAS in response to *B. cinerea* by being DE.

These results strongly indicate there are cis-elements associated with *B. cinerea* mediated DAS RI, and have identified some candidate *trans* elements for investigation to determine if they are involved in the *B. cinerea* mediated DAS. This is further supported by three of the putative *B. cinerea*-mediated DASed motif being enriched in validated RI events. Further work could involve binding experiments to determine if the cis and *trans* elements can bind. To further identify *B. cinerea*-

mediated RI motifs that bind to RBP, PIP-Seq or RNA immunoprecipitation coupled to sequencing (RIP-Seq) could be utilised in *B. cinerea* infected and non-infected Arabidopsis leaves. PIP-Seq uses a ribonuclease-based protein footprint sequencing approach to discover transcriptome wide interaction sites via identifying protein protected sites (PPS). Comparing the footprints from *B. cinerea* infected, and non-infected Arabidopsis leaves, could potentially identify PPS that only occur in infected leaves. Taking these sequences and utilising motif finding programs such as DEME or HOMER has the potential to identify *B. cinerea* specific motifs that could bind RBP. RIP-Seq could also be used to help determine if known RBP bind to different RNA sequences under the two conditions. Again, utilising motif finding programs to identify motifs that are novel to *B. cinerea* infected leaves. The disadvantage of RIP-Seq is that only one known RBP can be investigated at once. PIP-Seq has the advantage that it considers the combinatorial binding and regulation of multiple RBPs.

3.5 Conclusion

In conclusion, although the Cooke (2013) dataset has minimal time-points and no biological replicates, the fact that it compares favourably to the Windram *et al* (2012) dataset means it is a useful tool to identify some potential *B. cinerea* mediated DASed marker genes. Further validation of these genes has produced a list of seven *B. cinerea*-mediated DASed marker genes that can be used to evaluate RNA-Seq analysis techniques in other datasets. In addition, RT-PCR primers have been developed that can detect DAS of intron one of *AT4G39270* which could be utilised to determine if loss-of-function spliceosomal mutants affect the *B. cinerea*-mediated DAS of this gene.

Investigations into *B. cinerea*-mediated RI events has identified putative SREs within introns and in some cases identified potential *trans* elements that could be investigated to determine if they are involved in *B. cinerea*-mediated DAS. The motifs enriched in *B. cinerea*-mediated RI events in some cases show sequence

similarities to those seen in mammalian species adding further weight to the proposition that SRE can be evolutionarily conserved across a wide range of species. At the individual gene level an LRR-RLK gene *AT4G39270* has been identified to undergo *B. cinerea*-mediated DAS and been linked to a defence phenotype that is potentially specific to necrotrophic pathogens. Further work involving silencing of one spliceform could help to determine if one or both spliceforms are important for this defence response. Additional defence phenotype screening including other fungal and necrotrophic organisms would help to elucidate how specific the genes role is in the plant defence response.

The main work that is required is to investigate global DAS of Arabidopsis by *B. cinerea* with multiple biological replicates, in order for conclusions about genome-wide *B. cinerea*-mediated DAS to be drawn at the plant population level. This is carried out in "Chapter 4: Genome-wide analysis of Botrytis cinerea-mediated differential alternative splicing in Arabidopsis leaves".

4 Chapter 4: Genome-wide analysis of Botrytis cinerea-mediated differential alternative splicing in Arabidopsis leaves

4.1 Introduction

RNA-Seq has proven to be a powerful tool for detecting and investigating AS in *Arabidopsis* as well as profiling gene expression (Bhargava *et al*., 2013; Ding *et al*., 2014; Filichkin *et al*., 2010; Gulledge *et al*., 2012; Howard *et al*., 2013; Marquez *et al*., 2012). It has been used to investigate AS patterns in *Arabidopsis* under both normal conditions and whilst experiencing abiotic and biotic stress (Ding *et al*., 2014; Filichkin *et al*., 2010; Howard *et al*., 2013; Marquez *et al*., 2012). Here we concentrate on the application of RNA-Seq techniques to investigate alternative and differential splicing of genes in response to biotic stress in particular to the necrotrophic pathogen *B. cinerea*. In Chapter 3 we determined that the Cooke data set was relatively robust. However Cooke (2013) only investigated one biological replicate (two technical) per time point, meaning that this data cannot inform us about the biological variation of differential alternative splicing (DAS) and differentially expressed genes (DEG) (Cooke, 2013). Because both gene expression mechanisms and the splicing machinery are thought to be stochastic in nature, biological replicates are required if conclusions are to be able to be generalised to the plant population level (Melamud & Moult, 2009; Raj & van Oudenaarden, 2008). From the Cooke (2013) dataset we can see that a significant amount of DAS occurs in response to *B. cinerea* infection 24 hpi, and Windram *et al* 2012 determined that at this time point the majority of DEG have started to be differentially expressed. Thus, this time point was chosen to further investigate the genome-wide effect of *B. cinerea* infection on the *Arabidopsis* transcriptome

4.1.1 RNA-Seq analysis tools

RNA-Seq is a high throughput technology that sequences cDNA enabling comprehensive and quantitative analysis of the transcriptome. There are many

different pipelines that can be used to analyse RNA-Seq data that usually consist of the following stages, alignment, quantification, and DEG/DAS analysis.

4.1.1.1 Alignment tools

The mapping of the millions of reads produced from RNA-Seq to a reference genome or transcriptome is a critical task upon which any subsequent analyses are based. There are a multitude of mapping algorithms that align the vast number of reads to genomic or transcriptomic sequences, and determining which to use is a non-trivial decision. Aligners can be divided into two main categories, exon first and seed-and-extend, based on the type of algorithm used. The exon first aligners use unspliced aligners to map reads to exons producing read-clusters, then those reads that are unmapped are cut into smaller sequences, extended and used to find connections between the read clusters; seed-and-extend aligners divide reads into shorter subreads or kmers that are matched to the reference genome or transcriptome. These matches are then extended using a variety of algorithms and splice sites located (Figure 38). A subset of the seed-and-extend category are the multi-seed aligners; here multiple subreads are produced from each read and mapping location of the read is determined by the largest consensus set of subreads.

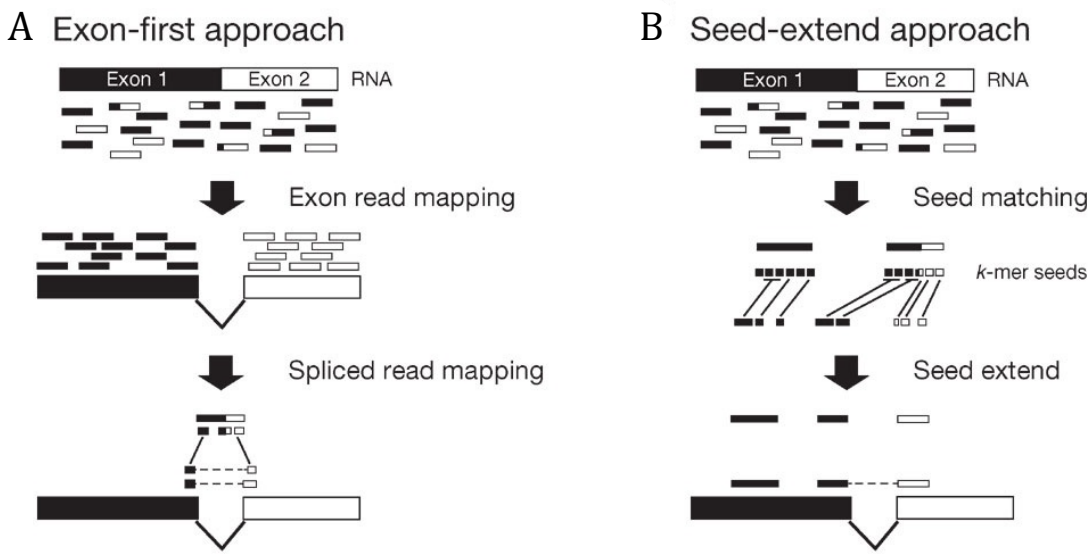


Figure 38 Illustration of the exon first and seed-and-extend algorithms. Exon first and seed-and-extend are the two main categories of spliced aligner algorithms. A. In the exon first algorithms unspliced aligners are used to map full length unspliced reads creating read-clusters, then the unmapped reads are split into shorter segments and mapped independently. The genomic regions surrounding the mapped read segments are then searched for possible connections between read clusters. B. In seed-and-extend all reads are divided into shorter subreads or kmers, these are then mapped to the genome (or transcriptome) and extended into larger alignments which may include gaps flanked by splice sites. Illustration by Garber *et al* (2011).

A literature review comparing eleven popular aligners from the three algorithm categories, exon first, seed-and-extend and multi-seed, was carried out to determine which aligners to use for analysis of our RNA-Seq data (Table 12).

Table 12. Comparison of popular RNA-Seq aligners. Column one gives details of the aligner. The eleven popular aligners compared can be divided into three categories based on the type of algorithm, exon first, seed-and-extend and multi-seed, column two. Column three briefly outlines the methods used to detect splice sites and column four details their relative performance. Column five is the number of times the paper detailing the aligner has been cited (data collected from Google Scholar on September 15th 2015).

Algorithm	Type	Splice site prediction method	Performance	Cited
Tophat, Trapnell *et al* 2009 (Trapnell, Pachter, & Salzberg, 2009)	Exon first	1. Aligns full reads via Bowtie 2. Searches for canonical (GT-AG) exon/intron boundaries 3. Seed & extend method used to fit unmapped reads over putative splice sites.	Biased towards canonical splice sites (Bryant, Shen, Priest, Wong, & Mockler, 2010).	3424
Tophat2, Kim *et al* 2013 (D. Kim *et al*., 2013)	Exon first	1. Aligns full reads via Bowtie1 or Bowtie2 2. Searches for canonical (GT-AG) exon/intron boundaries 3. Seed & extend method used to fit unmapped reads over candidate splice sites 4. Candidate splice sites used to correctly align multiexon-spanning reads	More accurately mapped paired end reads than GSNAP, RUM, MapSplice or STAR. Significantly faster than Tophat but still 17 times slower than STAR, although STAR requires 5 times more memory (in GB) (D. Kim *et al*., 2013). Biased towards canonical splice sites (Bryant *et al*., 2010).	859
SpliceMap, Au *et al* 2010 (Au, Jiang, Lin, Xing, & Wong, 2010)	Exon first	1. Split reads into 2 halves 2. Mapped to genome (aligner of choice) 3. Reads spanning splice junction have to have >half mapped to exon. 4. Seed & extend to find canonical splice site 5. Matched to hit on other side of intron	Biased towards canonical splice site (Bryant *et al*., 2010). Less sensitive and specific than Tophat, however this is possibly due to aligner choice (Grant *et al*., 2011).	232
MapSplice, Wang *et al* 2010 (K. Wang *et al*., 2010)	Exon first	1. Reads split into smaller 20-25bp tags 2. Mapped to genome with aligner to give candidate alignments 3. If not aligned but tag upstream and downstream are extended inwards to align middle unaligned tag	Detects canonical and non-canonical splice sites and is more sensitive and specific than Tophat & SpliceMap (Grant *et al*., 2011; K. Wang *et al*., 2010)	288
SOAPsplice, Huang *et al* 2011 (S. Huang *et al*., 2011)	Exon first	1. Maps full reads to genome 2. Unmapped divided into 2 segments 3. Longest segment at 5’	Higher detection and lower false positives than Tophat, MapSplice and SpliceMap on simulated data when read	50

		end mapped to reference 4. Searches for intron boundary in order of 'GT-AG' 'GC-AG' 'AT-AC'5. Paired end reads used to filter false positives.	coverage low (S. Huang *et al*., 2011).	
RUM, Grant *et al* 2011 (Grant *et al*., 2011)	Exon first	1. Bowtie maps read to genome & transcriptome 2. Merged with preference for transcriptome 3. Unmapped aligned to genome with BLAT 4. Merged into final alignment 5. Junction detect by read gap (default >15bp) & known splice signals	On simulated data where the alignments were known it has the second best accuracy and combined lowest false positive & false negative rate (Grant *et al*., 2011). About half run time of GSNAP but requires a large amount of memory (D. Kim *et al*., 2013).	140
GSNAP, Wu and Nacu 2010 (T. D. Wu & Nacu, 2010)	Seed-and-extend	1. Probabilistic model scores donor/acceptor sites 2. User submitted data of known splice boundaries for exon/introns 3. Filters out false positives using maximum entropy model	Cannot infer new splice sites but can suggest new exon combinations. More sensitive and specific than Tophat, simulated data best accuracy and combined lowest false positive and false negative rate (Grant *et al*., 2011).	607
QPALMA, DeBona *et al* 2008 (De Bona, Ossowski, Schneeberger, & Rätsch, 2008)	Seed-and-extend	Based on training set of previously identified splice sites	Biased to detecting type of splice site previously found, thus limits detection of novel splice sites plus it is biased towards canonical splice sites (Bryant *et al*., 2010).	139
STAR, Dobin *et al* 2013 (Dobin *et al*., 2013)	Seed-and-extend	1. Seed searching 2. Clustering/stitching/scoring	Compares favourably to aligners GSNAP, Tophat and MapSplice (Engström *et al*., 2013). 180 and 17 times faster than GSNAP and Tophat2 respectively (Dobin *et al*., 2013; D. Kim *et al*., 2013).Requires more GBs of memory than other aligners with exception of RUM (D. Kim *et al*., 2013).Not biased towards canonical splice sites (Dobin *et al*., 2013).	439
Subread/ Subjunc, Liao *et al* 2013 (Liao, Smyth, & Shi, 2013)	Multi-seed	1. Number of 16mer subreads (seed) produced for each read. 2. Subreads "vote" for best location to map read.	Less accurate than Tophat2 and STAR according to Magis *et al*, however the developers state it is 5% more accurate than Tophat2; it is faster than Tophat2 (Liao *et al*., 2013;	90

			Magis, Funk, & Price, 2015). Part of the statistical package R making it easy to incorporate into a pipeline (Liao, Smyth, & Shi, 2014).	
OLego, Wu *et al* 2013 (J. Wu, Anczuków, Krainer, Zhang, & Zhang, 2013)	Multi-seed	1. Burrows–Wheeler transform used in multiple steps to efficiently map seeds, locate junctions and identify small exons. 2. Statistical model used to score exon junctions by splice-site strength and intron size.	TopHat2 has better alignment precision and percentage of uniquely mapped reads, for both known and novel junctions (Gatto *et al*., 2014). Higher precision compared to STAR but lower sensitivity, slower than STAR and Tophat2 (D. Kim, Langmead, & Salzberg, 2015).	24

The aligners chosen to analyse RNA-Seq data in this thesis were Tophat2 and STAR based on their precision/sensitivity compared to other aligners and their ease and speed of use. Although STAR has large memory requirement this is negated by the fact that it is available to use via CyberVerse, a cyberinfrastructure for the life sciences (Goff *et al*., 2011). Once the aligner is chosen often there are quantification and DE algorithms that are designed to work alongside them, for example Cufflinks2 and CuffDiff2 for Tophat2, and FeatureCount for Subread. The Subread aligner is part of the Rsubread package available on Bioconductor, which can create a pipeline from alignment (using Subread) through to quantification (using FeatureCount). This pipeline will be compared to those created using the Tophat2 and STAR aligners.

4.1.1.2 Quantification of reads.

There are two main quantification mechanisms, the model based approach, for example Cufflinks and RSEM, which produce normalised read counts (B. Li & Dewey, 2011; Trapnell *et al*., 2010) and read count methods such as FeatureCount and htseq-count that produce raw count data (Anders, Pyl, & Huber, 2015; Liao *et al*., 2014). Normalised read counts are commonly expressed as fragments per kilobase per million mapped (FPKM) or transcripts per million (TPM). It has been argued that TPM should be the preferred choice due to it being independent of transcript length and having less variability between samples (B. Li & Dewey, 2011; Wagner, Kin, & Lynch, 2012). A recently developed tool which provides quantification in both TPM and FPKM that does not require mapping before quantification is Sailfish; this alignment-free k-mer-based approach disposes of the time-consuming step of aligning reads to a reference genome or transcriptome and has been shown to have better accuracy in determining transcript isoform abundances than Cufflinks (Kanitz *et al*., 2015; Patro, Mount, & Kingsford, 2014). An extensive review of 50 different pipelines carried out by Fonseca *et al* 2014 showed that the choice of tool used for quantification has the greatest impact of differences in detecting DEG between pipelines and that htseq-count consistently appeared in the top rankings regardless of the aligner used (Fonseca, Marioni, & Brazma, 2014). The quantification methods that we have decided to incorporate into our pipelines are Cufflinks, due to it being

developed to work alongside the Tophat2 aligner chosen, FeatureCount because it is part of the Rsubread package, and htseq-count because of its consistent high performance in the Fonseca *et al* 2014 pipeline evaluations, Sailfish will also be included due to its novel alignment-free k-mer-based approach.

4.1.1.3 Detecting differential expression of genes

There are many algorithms available for detecting DEG. Nookaew *et al* (2012) states that methods based on negative binomial modelling, such as DESeq, edgeR, and baySeq have improved specificity and sensitivities as well as good control of false positive errors with comparable performance to other methods. Of these three algorithms Nookaew *et al* (2012) found that edgeR detected more novel DEG than the other methods indicating that it may have a higher false discovery rate. However Rapaport *et al* (2013) found that it had the second lowest false discovery rate (FDR) of the algorithms it tested regarding differential expression (DE) of genes. Of those methods tested by Rapaport *et al* (2013) (Cuffdiff, DESeq, edgeR, LimmaQN, Limmavoom, PoissionSeq and baySeq), they found that edgeR and DESeq had slightly better accuracy in detecting DEG. However Anders *et al* (2010) found that the conservativeness of edgeR varied between those genes that are lowly expressed and those that are highly expressed. Weakly expressed genes are over-represented whilst few highly expressed genes are deemed to be differentially expressed by edgeR, whereas DESeq has comparable sensitivity to edgeR whilst being more balanced over the dynamic range. Bayseq discovered only one false positive in a highly expressed gene (Rapaport *et al*., 2013), indicating that it may be more stringent than other methods and thus have an increased false negative rate. The methods chosen to investigate DEG are Cuffdiff2, because it is part of the Tophat pipeline, edgeR and DESeq due to their slightly increased accuracy in detecting DEG compared to other methods. Both algorithms were chosen in order to detect DEG that are both lowly and highly expressed. DESeq2 was also used to detect DEG because the developers observed that DESeq can be overly conservative and thus improved the algorithm producing DESeq2 (Love *et al*., 2014).

4.1.1.4 Detecting differential alternatively spliced genes.

Detection of differential splicing can be carried out at three levels, the transcript level, the event level or the splice junction level, with Cooke (2013) finding novel DAS genes from each level of analysis. Here we will investigate pipelines at each of these levels to help determine if multiple levels of analysis are required.

There are many different tools available that detect DAS, although a large proportion of these cannot account for biological variation because they do not work with replicates, such as MISO, ALEXA-Seq, JETTA, SpliceSeq, SOLAS and SplicingCompass and therefore these tools were not used (Aschoff *et al*., 2013; Griffith *et al*., 2010; Katz *et al*., 2010; Richard *et al*., 2010; Ryan *et al*., 2012; Seok *et al*., 2012).

4.1.1.4.1 Transcript level approach

To detect DAS at the transcript level, whole transcripts must be reconstructed from the short reads produced from RNA-Seq. This is a complex process and a recent assessment of 25 transcript reconstruction methods for RNA-Seq found that for most transcripts, automated methods failed to identify all constituent exons, and often failed to assemble the exons into complete isoforms (Steijger *et al*., 2013). The three methods that we have chosen to detect DEG, Cuffdiff2, edgeR and DESeq2 can be used to detect differential transcript usage (Anders *et al*., 2012; Robinson, McCarthy, & Smyth, 2009; Trapnell *et al*., 2013). These packages output information about which transcripts are differentially expressed between treatments. In order to infer if genes are undergoing DAS the ratios of differentially expressed transcripts will be investigated for each gene. This will enable us to differentiate between DEG, where all transcripts would be increased by the same amount (and therefore there would be no proportional change in response to treatment) and genes undergoing DAS which would show a proportional change in transcripts.

4.1.1.4.2 Event level approach

For event level analysis the AS events first need to be determined. For this we use two packages, ASTALAVISTA and SUPPA, to detect the AS events (Alamancos *et al*., 2015; Foissac & Sammeth, 2007). The relative inclusion ratio and percentage spliced

in (PSI) can be used to detect differences in the proportions of AS events between treatments (E. T. Wang *et al*., 2008). The relative inclusion ratio is the ratio of the number of inclusion reads (i.e. reads in regions supporting the inclusion event) to the sum of inclusion plus exclusion reads, for each event. The percentage spliced in (PSI), which is the fraction of mRNAs that contain, for example, a spliced exon or part of an exon, or a retained intron, can be estimated as the ratio of the density of inclusion reads (the reads per position that support the inclusion isoform) divided by the sum of the densities of the inclusion and exclusion reads (E. T. Wang *et al*., 2008). Alamancos *et al* (2015) use a version of this to calculate the PSI of an event, which they define as the ratio of the abundances of the inclusion reads (given in transcripts per million (TPM)) to the abundances of the inclusion and exclusion reads (Figure 39).

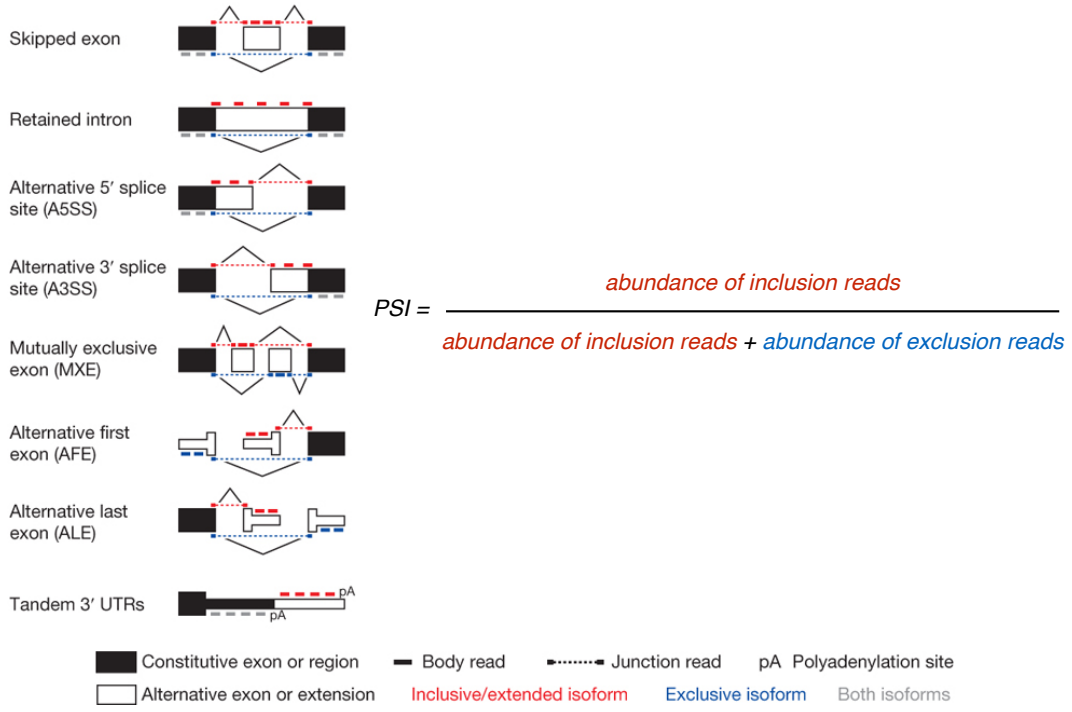


Figure 39. Percentage spliced in (PSI) is used to detect the ratios of AS events. The PSI of an event is calculated as the ratio of the abundance of inclusion reads (blue) over the abundance of the inclusion reads plus the abundance of exclusion reads (red). Figure adapted from Wang *et al* 2008 (E. T. Wang *et al*., 2008).

4.1.1.4.3 Splice Junction level approach

To determine differential splice junction (SJ) usage we used the junction files from Tophat2 and STAR to provide a count of reads spanning the SJ and determined

differential usage using Limma, a well-established package for the analysis of gene expression data arising from microarray and RNA-Seq data (Ritchie *et al*., 2015). To account for the differences in the type of data from microarrays (intensities that are essentially continuous numerical measurements) and RNA-Seq (integer counts) the voom function is used to estimate the mean-variance relationship robustly and non-parametrically from the data. Thus, downstream analysis is based on the assumption that data is of normal distribution rather than negative binomial (as is the case for EdgeR) with the main advantage of this being that the Limma pipeline provides accurate type I error rate control even when the number of RNA-Seq samples are small (Law *et al*., 2014).

4.1.2 Arabidopsis and DAS of known gene families in response to infection.

In this chapter, two main gene families will be investigated in detail, the genes that encode the JASMONATE ZIM-DOMAIN (JAZ) proteins, and the leucine-rich repeat receptor like kinases (LRR-RLKs). These gene families were chosen because they contain known *B. cinerea* defence responsive genes, but are involved in different parts of the defence pathway. LRR-RLKs are involved in PAMP/DAMP perception, whereas JAZs are involved in regulatory mechanisms of defence responsiveness because they are a crucial part of the Jasmonic acid (JA) signalling pathway.

4.1.2.1 The LRR-RLKs and alternative splicing

LRR-RLKs are one of the large subfamily of the Receptor Like Kinases (RLK), which have been shown to detect PAMPs (Mastrangelo, Marone, Laidò, De Leonardis, & De Vita, 2012; Shiu & Bleecker, 2001). RLKs are known to be regulated in part by AS, with a large proportion of LRR-RLKs shown to undergo AS in Arabidopsis (Gou *et al*., 2010; Mastrangelo *et al*., 2012). In Chapter 3 we have shown that the LRR-RLK *AT4G39270* undergoes DAS in response to *B. cinerea* infection, and identified nine other LRR-RLKs from the Cooke (2013) dataset that underwent DAS. Thus the LRR-RLK are known to be involved in the defence response and to undergo DAS in response to *B. cinerea*, thus making them a good gene family to investigate further.

4.1.2.2 The JAZs and alternative splicing.

The JAZ proteins are core components of the JA signalling cascade where they act as transcriptional repressors for JA-responsive target genes. Stress-induced production of JA causes the formation of COI1-JAZ co-receptor complexes targeting JAZ proteins for SCF^{COI1} mediated ubiquitination, thus releasing transcription factors from inhibition leading to activation of JA-responsive genes (Chini *et al*., 2007; Pauwels *et al*., 2010; Thines *et al*., 2007; Yan *et al*., 2007). Most *JAZ* genes contain an evolutionarily conserved intron that divides the Jas motif into a 20 N-terminal and a seven C-terminal (X5PY) amino acid submotifs. AS can result in retention of this intron producing truncated proteins with reduced ability to form stable complexes with COI1 in the presence of JA-Ile, resulting in some truncated splice variants being dominant repressors of JA signalling (H. S. Chung *et al*., 2010). An alternative splice variant of *JAZ10* (*JAZ10.4*) that lacks the Jas motif has been shown to mediate desensitisation to JA (Moreno *et al*., 2013). Thus, potentially AS may play a role in the defence response to pathogens. Investigating whether the proportion of AS transcripts alters in response to *B. cinerea* infection will help us to understand if AS plays a functional role in the plant defence of necrotrophic pathogens in regards to the JA signalling pathway. The fact that the JAZ family is known to be involved in the defence response and AS is shown to play a functional role in this response makes it an ideal family to study the effects of DAS in response to *B. cinerea*.

4.2 Aims and Objectives

The overarching aim of this chapter is to investigate the effect of pathogen infection on the transcriptome of *Arabidopsis*, with a focus on its impact on the alternative splicing landscape. In order to achieve this the following research objectives will be met:

3. Determine that infection is progressing as anticipated in the dataset by comparing differentially expressed genes (DEG) in this dataset to known Arabidopsis marker genes for *B. cinerea* infection
1. Evaluate different RNA-Seq pipelines, utilising differentially expressed marker genes for DEG pipelines, and the DAS marker genes detected in Chapter 3 in order to obtain a set of *B. cinerea*-mediated DAS genes.
2. Investigate this DAS gene set to
    a. Elucidate how extensive *B. cinerea*-mediated DAS is, and how it relates to the production of transcripts
    b. Identify putative regulatory mechanisms for DAS in the Arabidopsis defence response to *B. cinerea*
    c. Investigate two gene families known to be important for the *B. cinerea* defence response, the RLKs and JAZ proteins
    d. Determine how specific the defence related DAS is by comparing *B. cinerea*-mediated DAS genes to genes determined to be DAS in
        i. an abiotic stress, salt (Ding *et al*., 2014)
        ii. abiotic stress in a different type of organism, *Pseudomonas* (Howard *et al*., 2013)

4.3 Results

Arabidopsis leaves (leaf 7) from four-week-old plants were either inoculated with a suspension of *B. cinerea* spores or mock-inoculated, and harvested 24 hpi before visual lesions formed (Figure 40).

Figure 40. Plant material for RNA-Seq. Three biological replicates of mock and *B. cinerea* inoculated *Arabidopsis* leaves (leaf 7) were harvested 24 hpi before visual lesions were formed.

4.3.1 Quality analysis of RNA-Seq data

RNA-Seq was performed on RNA extracted from whole leaves (one leaf per sample) with between 25-49 million sequenced paired end reads (100 bp in length) per sample being obtained using the Illumina High-Seq sequencing system (Table 13).

Table 13. Number of RNA-Seq reads obtained per sample.

Biological replicate	Mock	Infected
1	26,480,741	25,630,664
2	29,805,886	49,290,487
3	28,349,890	31,315,650

Quality assessment of the RNA-Seq raw reads was carried out with FastQC on CyVerse (Goff *et al*., 2011). Distributions for the FASTQ quality scores for each sample were within acceptable ranges. The sequence call quality of the samples are good. In all samples the quality score was above the accepted cut-off of 28 indicating that trimming of reads is not required (Figure 41).

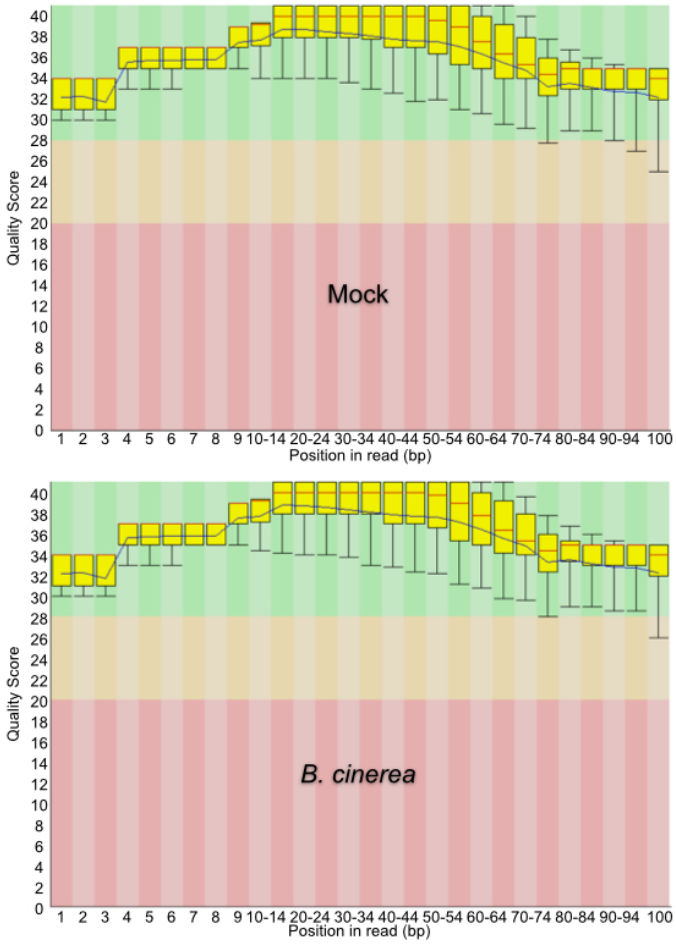


Figure 41. Quality of RNA-Seq raw reads. The sequence call quality of the samples is good as can be seen by the majority of the data falling within the "good" quality score area (the green area of the graphs). Two representative graphs are shown, mock top and *B. cinerea* infected bottom.

Quality assessment of our data with FastQC flagged warnings for Sequence Duplication Levels, Per base sequence content and Per base GC content. Upon visual inspection of the graphs these warnings were determined to be within expected parameters for RNA-Seq experiments, thus no further action was required (Figure 42). FastQC considers sequence duplication by creating a database of the first 200,000 reads input, and then it proceeds to check how many of the remaining reads are duplicates, producing a number relative to the original database. FastQC was originally written for genome sequencing where the expected Sequence Duplication Level graph should drop quickly to zero. Due to uneven expression, transcriptome data contains many duplicate reads, and thus an increase in the '10+' category

would occur. This is what is responsible for the warning flag in our data and thus this warning can be ignored (Figure 42A and B).

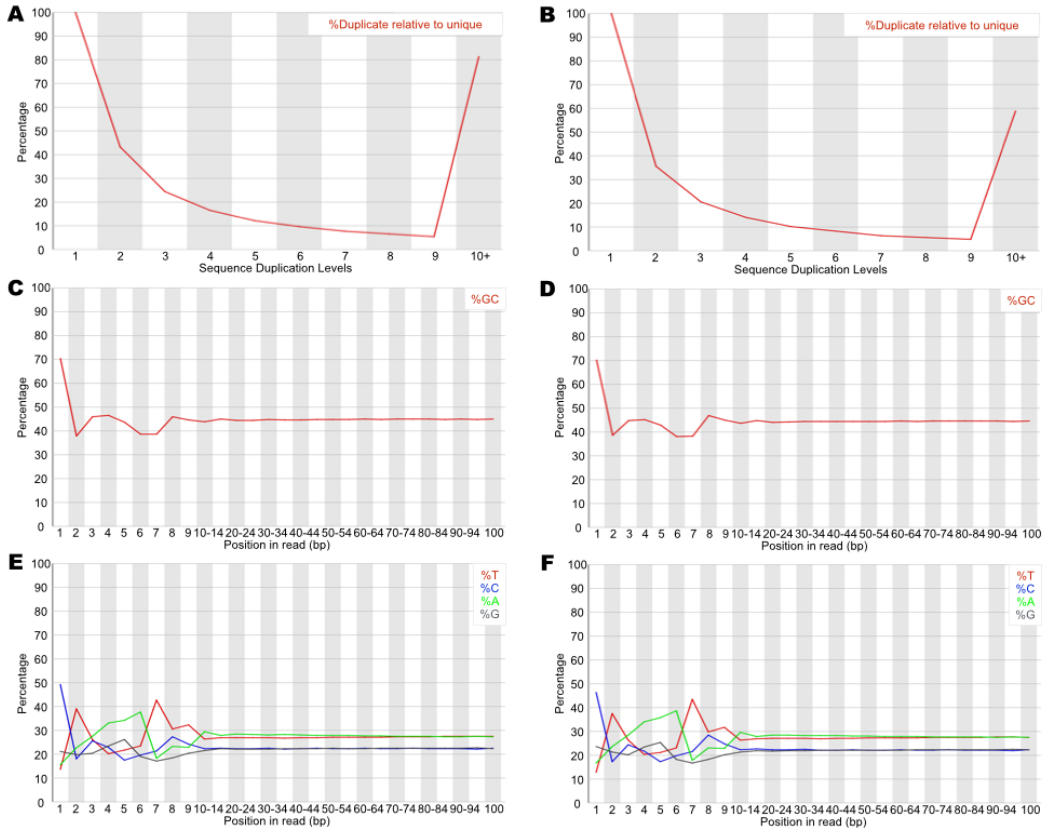


Figure 42. FASTQC graphs indicate no pre-processing of reads is required. Quality assessment using FASTQC produced three warnings, each of which can be explained by a visual inspection of the graphs resulting in no pre-processing of reads being required. A-B. The graph indicates that the Sequence Duplication Level warning is due to the increase in duplication levels in the 10+ category. This is expected with transcriptome data and thus does not require any adjustment. C-F. The graphs indicate that the warnings due to Per base sequence content (C and D) and Per base GC content (E and F) are due to bias in the early base pair (bp) positions. This can be ascribed to the library generation being based on the ligation of random hexamers and does not require any adjustment. Two representative graphs are shown per warning, mock left and *B. cinerea* infected right.

The two other warnings were for Per base sequence content and Per base GC content. Ideally there should be no variation along the length of the reads, however, if library generation is based on the ligation of random hexamers it often results in primers showing bias during the first 12 bp causing a warning to occur. This is what has caused the warning flags in our data, and as this is anticipated with RNA-Seq data no pre-processing of reads was required (Figure 42C-5F).

4.3.2 Evaluating the analysis methods

The alignment of reads is the first and potentially one of the most crucial parts of the analysis pipeline because all downstream analysis can be affected by it.

4.3.2.1 Alignment of reads

Of the three aligners investigated, Subread had the highest percentage of reads aligned to the Arabidopsis genome, followed by Tophat2, and then STAR. As expected mock samples had a higher percentage of reads aligned than infected samples. This is because infected samples will also contain reads originating from *B. cinerea* that should not align to the Arabidopsis genome (Table 14). The percentage of reads aligned by Subread was higher than expected possibly indicating that some of these reads are incorrectly aligned. Subread alignment was run with default settings. The maximum number of mismatched bases by default is 3. Potentially this could enable false positives to occour, accounting for the higher than expected alignment levels seen.

Table 14. Percentages of reads aligned by three different aligners. Subread aligned the largest percentage of reads followed by Tophat2 then STAR.

	Tophat2	STAR	Subread
Mock 1	95.4%	92.59%	100%
Mock 2	95.7%	91.34%	99.9%
Mock 3	95.0%	89.33%	99.5%
Infected 1	79.9%	79.54%	91.0%
Infected 2	88.2%	86.27%	93.4%
Infected 3	93%	93.47%	96.4%

4.3.3 Multiple pipelines are utilised for detecting differential alternative splicing and differentially expressed genes

For both DAS and DEG analysis, multiple pipelines were investigated. Utilising multiple detection analysis has been shown to improve detection rates when investigating DEG and is likely to have the same impact on DAS (Y. Guo *et al*., 2014).

It can reduce false positive (by excluding those genes only detected to be DE or undergo DAS in one pipeline) whilst including those that would be missed by utilising only one method (Zappia *et al*., 2015). However, using multiple algorithms can take considerable computational and analytical time. Determining at which point the law of diminishing returns occurs or if there are any algorithms consistently poorly performing could help to narrow down which pipelines or methods should be included or excluded in further analysis.

4.3.3.1 CuffDiff2 pipelines are the best at detecting DEG

Twenty different pipelines were used to detect DEGs. These consisting of three aligners (Tophat2, STAR and Subread), combined with two different quantification methods (HTSeq and FeatureCount) plus the alignment-free quantification method Sailfish, and four algorithms that detect DEG (Cuffdiff2, edgeR, DESeq and DESeq2) (Table 15).

Table 15. Pipelines for determining DEG. The pipelines consist of different combination of three aligners (column 2), three quantification methods (column 3), and four algorithms to detect DEG (column 4).

Pipeline	Aligner	Quantifier	DEG algorithm
1	STAR	-	CuffDiff2
2	STAR	FeatureCount	DESeq
3	STAR	FeatureCount	DESeq2
4	STAR	FeatureCount	edgeR
5	STAR	HTseq	DESeq
6	STAR	HTseq	DESeq2
7	STAR	HTseq	edgeR
8	Subread	FeatureCount	DESeq
9	Subread	FeatureCount	DESeq2
10	Subread	FeatureCount	edgeR
11	Tophat2	-	CuffDiff2
12	Tophat2	FeatureCount	DESeq
13	Tophat2	FeatureCount	DESeq2
14	Tophat2	FeatureCount	edgeR
15	Tophat2	HTseq	DESeq
16	Tophat2	HTseq	DESeq2
17	Tophat2	HTseq	edgeR
18	-	Sailfish	DESeq
19	-	Sailfish	DESeq2
20	-	Sailfish	edgeR

To compare the 20 pipelines and determine that the *B. cinerea* infection is progressing as anticipated, 41 marker genes from the list of genes previously identified as differentially expressed (DE) during *B. cinerea* infection were used (Windram *et al*., 2012). Except for those pipelines containing DESeq, the pipelines appear to have remarkably similar outputs, performing roughly equally in detecting the DE marker genes (Figure 43).

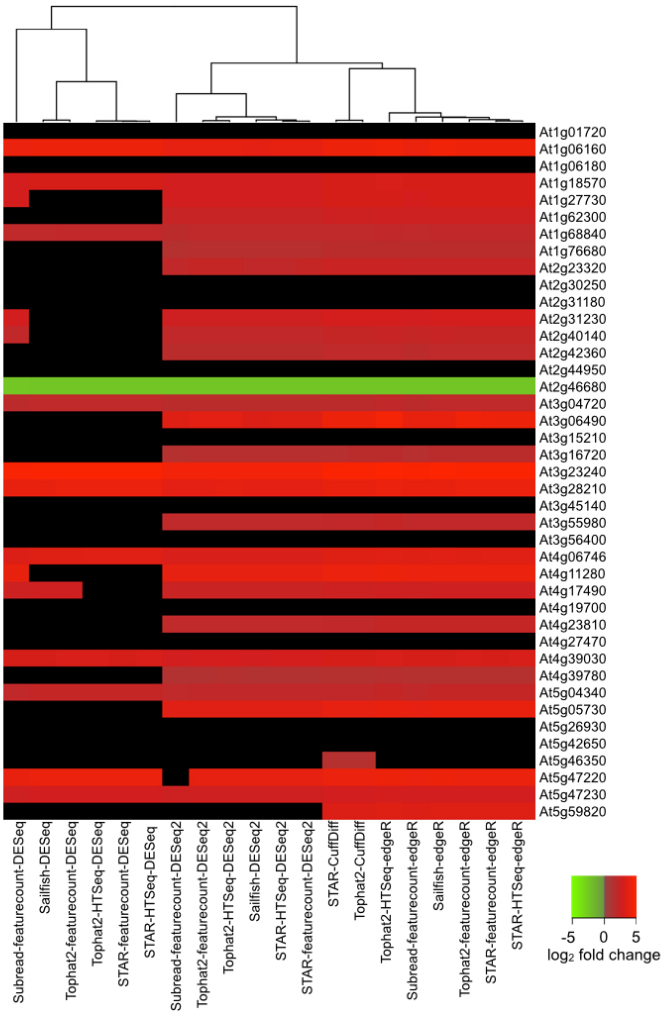


Figure 43. Heatmap illustrating differential expression of the 41 marker genes as detected by our 20 pipelines. The pattern of differentially expressed genes during *B. cinerea* infection from our pipelines broadly agrees with previously published data except for DESeq, which appears to be too conservative. Black bars represent genes that were not detected to be DE. Varying shades of red represent statistically significantly up-regulated genes at the 5% FDR with brighter red having a higher fold change, whilst varying shades of green represent statistically significantly down-regulated genes at the 5% FDR with brighter green having a higher fold change.

Hierarchical clustering used to produce the dendrogram at the top of (Figure 43) shows that the DE algorithm utilised appears to have the greatest impact on detecting DEG compared to other components of the pipelines. This is ascertained because the main similarity in each clade is the DE algorithm used.

Only one of the key marker genes was down-regulated in response to *B. cinerea*. To further investigate down-regulated genes we compared DE of all genes in TAIR10 that were deemed differentially expressed in any pipeline. Interestingly the method

used to detect DEG also affected the proportion of DEG that were deemed to be down and up-regulated. For Cuffdiff2 the proportion of DEG that were up-regulated was approximately 76%, 24% were down-regulated, for edgeR approximately 80% of DEG were up-regulated the remaining 20% were down-regulated, whereas DESeq2 had approximately 67% of DEG deemed to be up-regulated with 33% being down-regulated, DESeq detected 78% up-regulated genes (Figure 44).

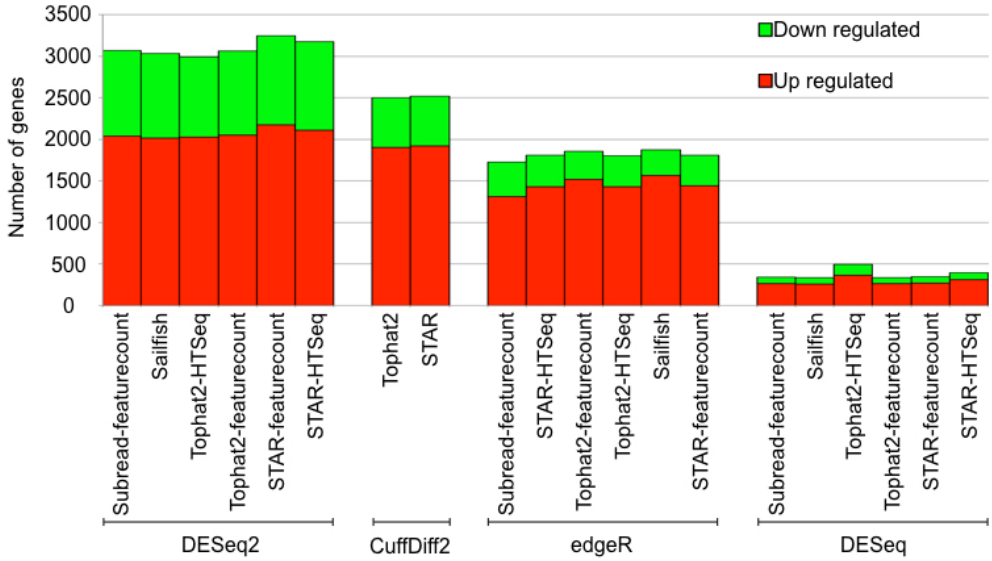


Figure 44. The algorithm used to determine differential expression of genes affects the proportion of down and up-regulated genes detected. The DESeq2 pipelines detect the largest number of down-regulated genes overall. However as a proportion of total number of DEG edgeR detects the largest number of down-regulated genes, the CuffDiff2 data represents the middle ground detecting fewer than DESeq2 but more than edgeR.

4.3.3.2 Expected DEG are detected in our dataset

The majority of the pipelines (excluding DESeq pipelines) detected 27 of the 41 marker genes. However the gene *AT5G59820* was not detected by any of the DESeq2 pipelines, and *AT5G46350* was only detected by the pipelines containing CuffDiff2 (Figure 43). Of the 27 genes detected to be DE in the majority of the pipelines, DESeq fails to detect ten of them in all of the pipelines where it is used, and four are only detected as DE in the Subread->FeatureCount->DESeq pipeline, indicating that DESeq is too conservative (Figure 43). The information gained from looking at the marker genes regarding our pipelines indicates that those pipelines containing DESeq are too conservative and the CuffDiff pipelines may be the most sensitive.

For eleven of the 12 DE marker genes not detected in our dataset their absence can be explained because they either have stochastic expressional profiles, randomness in translation and transcription can lead to cell-to-cell gene expression changes being stochastic in nature, or DE does not occur at 24 hpi. Six genes (*AT1G01720*, *AT2G30250*, *AT2G44950*, *AT3G45140*, *AT3G56400* and *AT5G26930*) appear to be stochastic in nature rather than DE because published literature give conflicting expression profiles. Four are not DE at 24 hpi in the Windram *et al* (2012) dataset (*AT1G06180*, *AT4G19700*, *AT4G27470* and *AT5G42650*). The time of first DE of *AT2G31180* the TOFDE was 24 hpi (Windram *et al*., 2012). Thus if our infection was progressing slightly slower it may also not yet be DE in our dataset.

This shows that DEG can be detected from our RNA-Seq data and that infection is progressing as expected, indicating that it is reliable enough to use for detecting DASed genes.

The DEG pipelines chosen to be used in future analysis are Tophat2->CuffDiff2 and STAR->CuffDiff2 because they are the only pipelines to detect DE of At5g46350. These pipelines also occupy the middle ground regarding percentage of up-regulated differentially expressed genes. In addition CuffDiff2 has the advantage that it is part of the tuxedo suite of Bioinformatics tools, which are widely used in the Bioinformatics community as well as being incorporated into a variety of cyber-infrastructures designed for biologists such as CyVerse and Galaxy (Blankenberg *et al*., 2010; Giardine *et al*., 2005; Goecks *et al*., 2010; Goff *et al*., 2011). This results in a list of 2577 genes differentially expressed in response to *B. cinerea* at 24hpi.

4.3.4 Differentially alternatively spliced genes can be detected at the transcript, event and splice junction level

Cooke (2013) detected DAS at three levels, the transcript, the event and the splice junction level, with each level detecting some DAS genes that are not detected in the other levels. Thus we compare pipelines for each level individually before combining the DASed genes from our chosen pipelines/methods into one dataset.

4.3.4.1 Transcript level

Each of the pipelines used to detect DEG can also be adapted to detect differentially expressed isoforms (DEI). Due to the poor performance of DESeq in detecting the DE marker genes, the DESeq pipelines were not used to detect DEI, resulting in 14 different pipelines being compared (Table 16).

Table 16. Pipelines for determining differentially expressed isoforms. Pipelines consist of different combinations of three aligners, three quantification methods, and three algorithms to detect DEI.

Pipeline	Aligner	Quantifier	DEI algorithm
1	STAR	-	CuffDiff2
2	STAR	featureCount	DESeq2
3	STAR	featureCount	edgeR
4	STAR	HTSeq	DESeq2
5	STAR	HTSeq	edgeR
6	Subread	featureCount	DESeq2
7	Subread	featureCount	edgeR
8	Tophat2	-	CuffDiff2
9	Tophat2	featureCount	DESeq2
10	Tophat2	featureCount	edgeR
11	Tophat2	HTSeq	DESeq2
12	Tophat2	HTSeq	edgeR
13	-	Sailfish	DESeq2
14	-	Sailfish	edgeR

The comprehensive, non-redundant Arabidopsis reference transcript dataset, AtRTD, constructed by Zhang *et al* (2015) was used in the estimation of transcript abundances using our 14 pipelines. AtRTD consists of 14677 multi isoform genes, 56% of genes (18948 genes) only had one isoform and thus were excluded from this analysis. From the 14677 multi-isoform genes, 2738 DEI were detected, originating from 2097 genes, in response to *B. cinerea* 24 hpi. Each of the pipelines detected some unique DEI. The method of differential detection appears to have the greatest influence (as it did with DEG). However the quantifier also appears to have a strong influence with the Sailfish->edgeR and Sailfish->DESeq2 pipelines being more like the cuffDiff2 pipelines than the other pipelines. This is ascertained because these pipelines are in the same clade in the dendrogram seen in Figure 45.

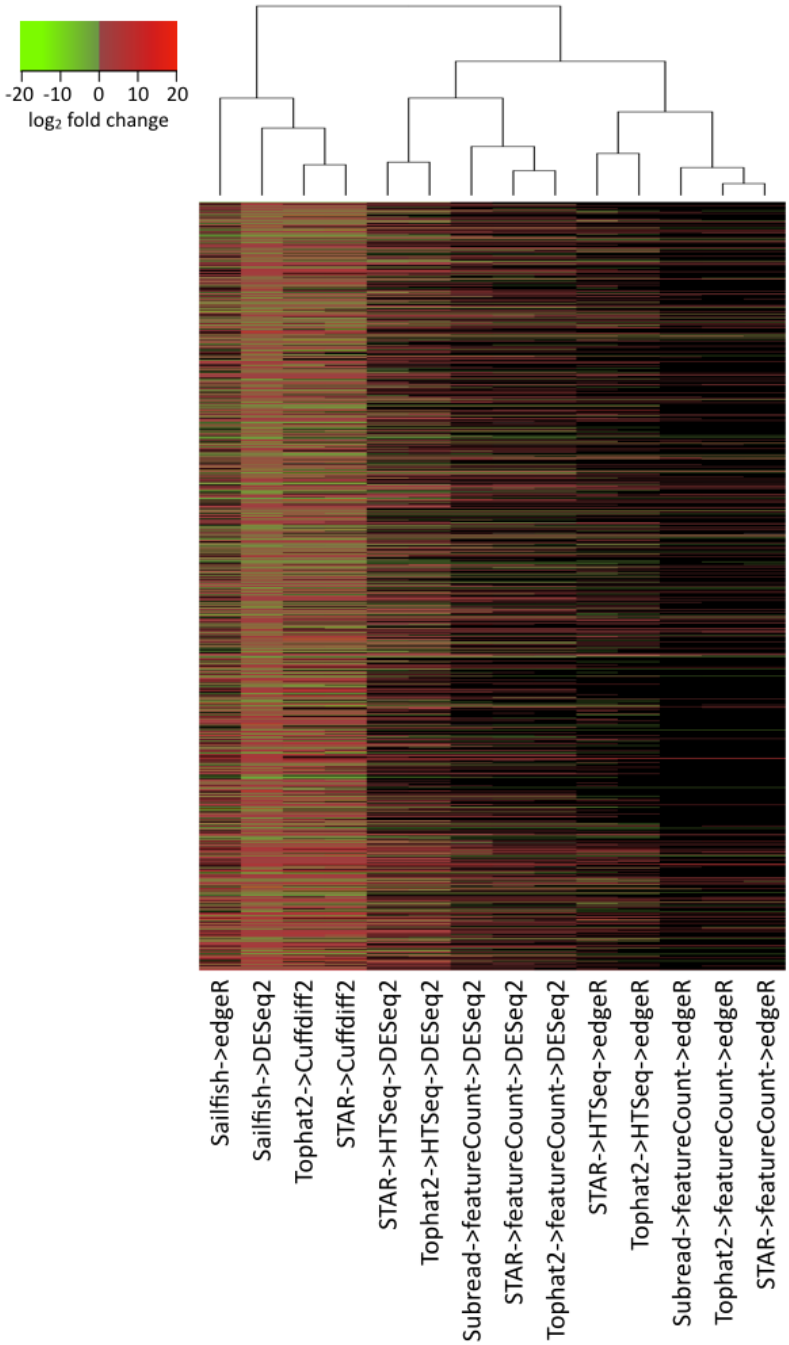


Figure 45. Heatmap illustrating the affect of pipelines on decting DEI. The dendrogram at the top of the heatmap illustrates the arrangement of clades produced by hierarchical clustering. Black bars represent genes that were not detected to be DE at the 5% FDR. Varying shades of red represents statistically significant up-regulated genes with brighter red having a higher fold change, whilst varying shades of green represents statistically significant down-regulated genes with brighter green having a higher fold change.

Why the DE algorithm used has a higher impact on fold change than the quantifier is unknown. However, each tool utilises a different normalisation and analysis technique to calculate the magnitude of fold change. DESeq2 and edgeR are both

count-based algorithms whilst CuffDiff2 uses a relative abundance method. Interestingly, when Sailfish is used as the quantifier these pipelines are more similar to the CuffDiff2 pipelines than the other pipelines containing edgeR or DESeq2. Sailfish outputs an estimated read count which the authors state can be used as an input for edgeR and DESeq2 (Patro *et al*., 2014). The estimated read count is the expected number of reads from each transcript based on the relative abundance estimate per transcript. Thus it appears that the difference in magnitude of the fold change of the isoforms is due to the different analysis techniques of count data verses relative abundances.

4.3.4.1.1 DEI methods detect genes that are differentially expressed at the both the gene and isoform level

DEI can arise from gene level regulatory mechanisms that would affect the overall expression of the gene, such as transcription factors and chromatin folding, or the regulatory mechanisms may effect individual or subgroups of transcripts by targeting alternative promoters or splicing factors (Howard *et al*., 2013). Of the 2738 DEI, 2277 isoforms (from 1895 unique genes) had at least one isoform in each gene that was not differentially expressed, potentially indicating that regulation of individual or sub groups of transcripts is occurring. Thus this subset of genes can be utilised to identify genes that undergo DAS.

4.3.4.1.2 Comparisons of DEI of previously validated genes that are DAS in response to *B. cinerea* 24hpi

Previously, we investigated 17 DAS events from eight different genes using HT RT-PCR. Five of these genes showed DAS in response to *B. cinerea* 24 hpi (*AT1G74590*, *AT4G20480*, *AT4G39270*, *AT5G16800* and *AT5G48657*), and thus can be used to help validate our pipelines. Only one of these, *AT5G48657*, was deemed to have DEI in response to *B. cinerea* 24hpi in the majority of our pipelines (all of the DEI pipelines except for the Subread -> featureCount -> edgeR pipeline). The only other marker gene detected to have DEIs is *AT4G39270*, where one of the pipelines, Sailfish->DESeq2, detected DEI at the 5% FDR. This could potentially indicate that

determining DAS solely at the transcript level results in a large number of false negatives, specifically if genes where all transcripts are DE are not included in the dataset of potential DAS genes. Further evidence for this comes from the gene *AT1G74590*, where the *B. cinerea*-mediated DAS has been validated at 24 hpi. *AT1G74590* has two isoforms, both of which are detected to be DEI in response to *B. cinerea*. Thus they were filtered out of the dataset of potential DAS genes. This indicates that to identify the maximum number of DAS genes, the complete list of DEI should be investigated because DAS and differential gene expression (DGE) can occur at the same time, however with the caveat that when all isoforms of a gene are DEI further analysis may be required to determine if they represent only DGE. For further investigations into DEI the two Cuffdiff2 pipelines will be utilised to give continuity with DEG analysis, plus the Sailfish->DESeq2 pipeline, because it is the only pipeline that detected *AT4G39270* to have a DEI. This gives a dataset of 2332 unique DEIs originating from 1846 genes.

To determine if DEI analysis provides useful information regarding DAS of genes that cannot be inferred from the other two other levels of analysis, event level and splice junction level analysis were also investigated and comparisons drawn.

4.3.5 Event level analysis

The event level analysis is concerned with identifying specific AS events that differ between mock and infected samples, rather than reconstructing the isoform and determining if there is a difference in isoforms between treatments as is the case with the DEI analysis. Here we use the PSI of an event to determine, in a pairwise manner, *B. cinerea*-mediated DAS events detecting skipped exon (SE), alternative 3′ splice site (A3SS), alternative 5′ splice site (A5SS), mutually exclusive exons (MX), retained introns (RI), alternative first exon (AFE) and alternative last exon (ALE).

4.3.5.1 SUPPA pipelines utilising Robinson and Smyth exact test do not detect expected DAS events

The SUPPA software identifies events from an annotation file (in our case the AtRTD file constructed by Zhang *et al* (2015)). It then determines the PSI value per sample for each of the events using transcript abundances. Three different methods were used to calculate the transcript abundances provided to SUPPA; the Sailfish program which provided TPM and is recommended by the developers of SUPPA, plus two other pipelines, Tophat2->Cufflinks2 and STAR->Cufflinks2. At the time of testing, the SUPPA software did not include significance testing to determine DAS between treatment, thus the Robinson and Smyth exact test (Robinson & Smyth, 2008) was performed in R with the Benjamini & Hochberg multiple testing correction method being applied to determine if there was a significant difference at the 5% FDR between the mock and infected samples. In line with previously published literature that utilised PSI to determine DAS only transcripts with a PSI ≥10% underwent testing for DAS (E. T. Wang *et al*., 2008).

The Sailfish->SUPPA pipeline detected the lowest number of DAS events; 708 events from 622 genes, and the Tophat2->Cufflinks->SUPPA detected the most, 859 events from 783 genes, with the STAR->Cufflinks->SUPPA pipeline detecting 820 DAS events from 745 genes. Although the number of events detected to undergo *B. cinerea*-mediated DAS in each pipeline is similar, there was little overlap between the events detected, with only 37 DAS events detected by all pipelines (Figure 46).

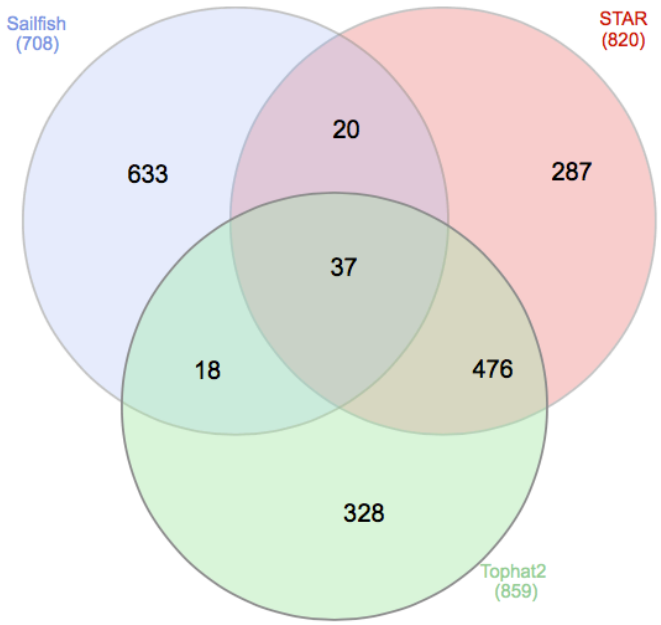


Figure 46. Venn diagram of DAS events detected by different algorithms. There is little overlap between the three SUPPA pipelines. Each of the pipelines utilised, Tophat2->Cufflinks2 (blue), STAR->Cufflinks2 (red) and Sailfish->SUPPA (green) detect many unique *B. cinerea*-mediated DAS events, with only 37 events being detected by all three pipelines.

None of the five previously validated DAS events are in the 1799 DAS events from 1467 genes that are detected by the SUPPA pipelines and only 1.8% (71 out of 1467) of the genes deemed to undergo DAS events are classified as DAS in response to *B. cinerea* in the Cooke (2013) dataset. This lack of overlap and the fact that the SUPPA event pipelines do not detect the validated DAS events may potentially indicate that these SUPPA pipelines may not be accurately detecting DAS events.

4.3.5.2 ASTALAVISTA pipelines utilising Robinson and Smyth exact test detect a few expected DAS events

To help determine if this was an issue with how SUPPA detected the AS events the ASTALAVISTA software was used to extract all events in the Zhang *et al* (2015) annotation file and for the four main AS events RI, A3SS, A5SS and SE the relative inclusion levels were determined, see section “2.2.4.2” in methods for details. Statistical significance was determined using the Robinson and Smyth exact test (Robinson & Smyth, 2008) with the Benjamini & Hochberg multiple testing correction method being applied at the 5% FDR to determine DAS events. This

method was implemented in two pipelines, one where the junction reads were determined using STAR and one where they were determined using Tophat2. In total the STAR pipeline detected 567 DAS events with the Tophat2 pipeline detecting 668 DAS events. The pipelines utilising STAR and Tophat2 detected 411 and 506 DASed RI events resulting from 362 and 434 genes respectively with 52% (314 events out of 603) being detected in both pipelines. In the Tophat2 pipeline only two out of the five previously validated genes, *AT1G74590* and *AT4G39270*, were detected to undergo DASed RI events in response to *B. cinerea* infection at the 5% FDR level, with only the first intron in *AT4G39270* being DASed (adjusted p value 6.2 X 10^{-5} and 0.006 respectively). The STAR pipeline only detected *AT1G74590* of the validated genes to be significantly DASed in response to *B. cinerea* (adjusted p value 1.9 X 10^{-5}). Relaxing the FDR to 10% did not result in more of the validated transcripts being detected.

4.3.5.3 SUPPA with diffSplice detects expected DAS events

The most recent version of SUPPA (2.1.0) has a DAS algorithm, diffSplice, incorporated into it where comparing the change in PSI (ΔPSI) between conditions with the distribution of the ΔPSI between replicates as a function of the gene expression (measured as the expression of the transcripts defining the events), statistical significance is determined (Alamancos *et al*., 2015). Utilising this statistical analysis two out of the five marker genes were statistically significantly DASed with a 5% FDR, *AT1G74590* and *AT5G16800* ($p<0.05$), with intron one of *AT4G39270* having a p value of 0.06, and *AT5G48657* had a p value of 0.09. Thus if we use the 10% FDR four out of the five marker genes are validated, indicating that SUPPA with diffSplice is a good algorithm for analysing DAS at the event level. Thus, only this dataset will be utilised for further event level analysis.

4.3.5.4 Retained introns are the most common event type

At the 10% FDR 8269 events originating from 5269 genes undergo *B. cinerea*-mediated DAS. The highest proportion of *B. cinerea*-mediated DAS event type is RI, with 53% of DAS events being of the RI type, this is only slightly higher percentage

than the number of RI AS events detected (49.8%) (Figure 47). In chapter 3 section 3.3.1.2 we detected an increase in the percentage of RI in response to infection in the Cooke (2013) dataset, whereas here no increase in the proportion of transcripts with RI is detected. This could be due to additional replicates reducing the noise in analysis or because we only investigating one time point, further work including time-series could help to determine which is the case.

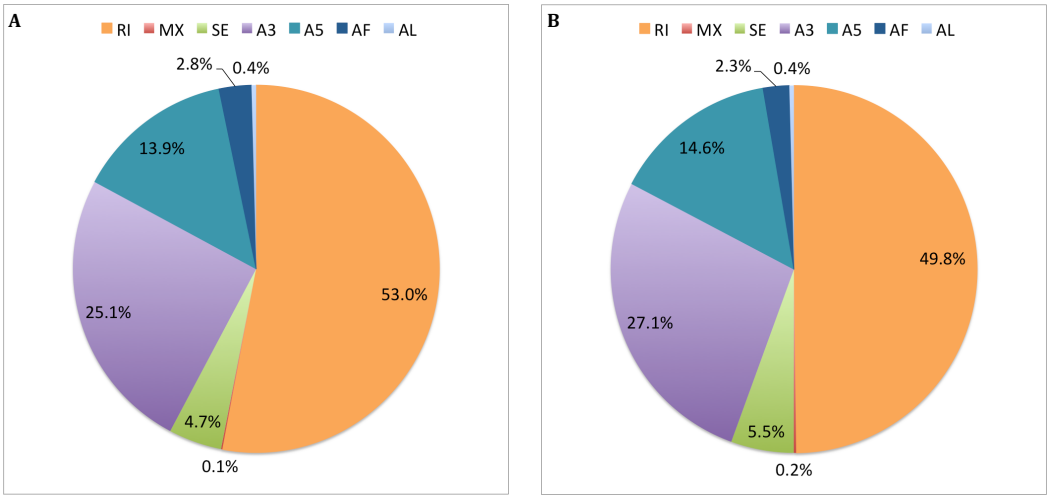


Figure 47. Distribution of event type. A. In *B. cinerea*-mediated DASed events. B. All alternatively spliced events detected.

Splice junctions (SJ) can be utilised differently depending upon condition, for example in different tissues or under differing environmental conditions. Here we investigate differential splice junction usage in response to *B. cinerea* 24 hpi.
The RNA-Seq reads are aligned to the AtRTD to identify exon-exon splice junctions utilising Tophat2 and STAR separately. The junction files from these aligners were used to count the number of reads spanning each SJ. This was then processed using Limma in R to determine statistically significant differential usage of SJs utilising the linear modelling approach implemented by lmFit and the empirical Bayes statistics implemented by eBayes. SJ that had less than 3 reads in all treatments were filtered out and the "voomWithQualityWeights" option was used to provide a more powerful analysis as recommended by Liu *et al* (2015). This works by helping to reduce the impact of the variation in sample quality. By down weighting

observations from more variable samples noise is reduced. Although completly removing the sample from the analysis would reduce the noise more efficiently this results in reducing the power of the analysis. Thus voomWithQualityWeights is a compromise, by partially reducing the noise of variable samples without removing them from the analysis which could negatively impact its power (Liu *et al* 2015).
In response to *B. cinerea* infection there are 11289 and 11557 SJs that display differential SJ usage between mock and *B. cinerea* infected Arabidopsis leaf samples (3 biological replicates per treatment) at the 5% FDR level in the STAR and Tophat2 pipelines respectively. There are more SJs that show an increase in usage in response to *B. cinerea* infection (7841 and 7913 for STAR and Tophat2 respectively) compared to those that have a decrease in usage (3448 and 3644 for STAR and Tophat2 respectively). However, decrease in usage may be harder to detect because it requires transcripts to degrade, thus this could be an artefact rather than of biologically significant (Figure 48).

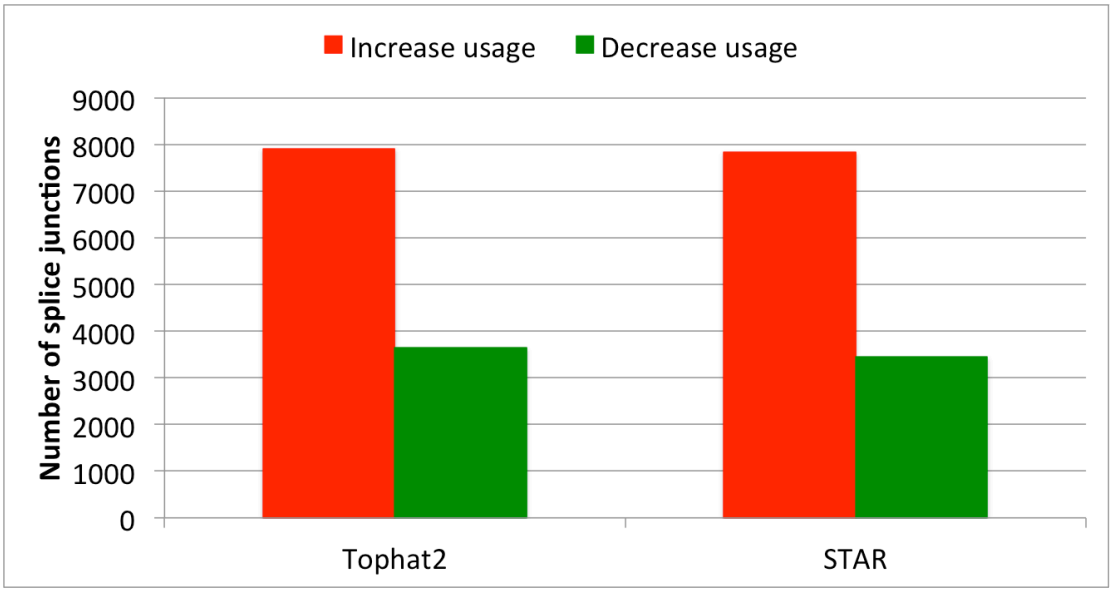


Figure 48. Differential splice junction usage between mock and *B. cinerea* infected Arabidopsis leaves. In both pipelines a greater number of SJs with increased usage (red bars) is detected in response to *B. cinerea* 24hpi than decreased usage (green bars). The Tophat2 pipeline detects slightly more differential SJ usage than the STAR pipeline regardless of increased or decreased usage.

The Tophat2 pipeline detected a larger number of statistically significant differential SJ usage than the STAR pipeline, although there is a large overlap between the two pipelines, 10451 SJ (84.3% of all SJ with differential usage) (Figure 49).

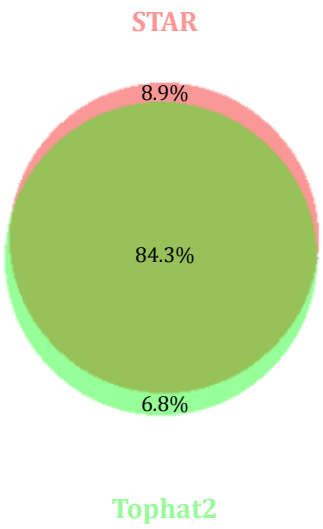


Figure 49. Venn diagram illustrating the large overlap between the two differential splice junction usage pipelines. Venn diagram created using Biovenn (Hulsen *et al*., 2008).

Both pipelines detect differential splice junction usage with similar fold changes; the average fold change is 1.6 for the Tophat2 pipeline and 1.57 for the STAR pipeline. To minimise the number of false positives, we will focus on the 10451 SJ that show differential usage in response to *B. cinerea* in both pipelines. These 10451 differential SJ usage events originate from 3396 genes. Of these only four genes contain SJs where some SJs are increased in usage whilst others are decreased in response to *B. cinerea* infection (*AT2G36380*, *AT3G56860*, *AT4G23895* and *AT5G58510*). When all SJs in a gene increase (or decrease) by approximately the same amount it could represent DE of the gene rather than DAS. Thus, these genes are excluded from our investigation in DAS and related motifs. Of those genes detected to have *B. cinerea*-mediated differential splice junction usage in both pipelines, where all possible splice junctions show differential usage and the difference between the minimum and maximum log-fold change of SJs usage is less than 0.5, 67% were also shown to be DEG in the section 4.3.3.1 compared to only 31% where either only a sub population of the splice junctions showed *B. cinerea*-mediated differential usage or the log fold change was greater than 0.5. Thus, to achieve a conservative dataset, those genes where all splice junctions showed *B. cinerea*-mediated differential usage or the maximum log-fold change of SJ usage is less than 0.5 were excluded from further analysis. Filtering on these criteria results in

8091 SJs that undergo *B. cinerea*-mediated differential usage from 2815 unique genes.

4.3.5.5 Comparisons with previously validated differentially alternatively spliced genes

Five of the nine SJ co-ordinates from the retained intron events (originating from five genes) that have been validated using HT RT-PCR displayed differential splice junction usage in response to *B. cinerea* 24hpi at the 10% FDR . At the 5% FDR level (*AT1G74590*, *AT4G39270* co-ordinates Chr4:1827868-18278857 and Chr4:18278985-18279066), with two being statistically significant at the 10% FDR level (*AT5G16800* Chr5:5524701-5524811 and Chr5:5524735-5524836). Two of the SJ co-ordinates have insufficient read coverage for analysis (*AT4G39270* co-ordinates Chr4:18278768-18278868 and Chr4:18279088-18279163). The SJ associated with *AT4G20480* did not show differential SJ usage (adjusted p value 0.281 and 0.236 for Tophat2 and STAR pipelines respectively). For the gene *AT5G48657* although differential splice junction usage was detected, it was detected in all SJs of the gene with only a 0.05 difference in the log fold change and had, therefore been filtered out. This could potentially indicate that our filtering criteria is too stringent.

4.3.6 Unique DAS genes are detected at the isoform event and splice junction level of analysis

Genes were detected to undergo *B. cinerea*-mediated DAS at the isoform level (using Tohat2->CuffDiff2, STAR->CuffDiff2 and Sailfish->DESeq2 pipelines), the event level (using SUPPA with diffSplice) and SJ level (the intersect of Tophat2->Limma and STAR->Limma). Combining the information from each level of analysis results in 7578 genes that are detected to undergo *B. cinerea*-mediated DAS at 24 hpi. Of these 7578 DAS genes, 17.6% and 7.19% are uniquely detected by the SJ and DEI pipelines respectively. The event level analysis detected the highest proportion of the 7578 *B. cinerea*-mediated DAS genes with 69.5% being detected at this level, with 49.82% of these 7578 genes being uniquely detected by the event level pipeline (Figure 50).

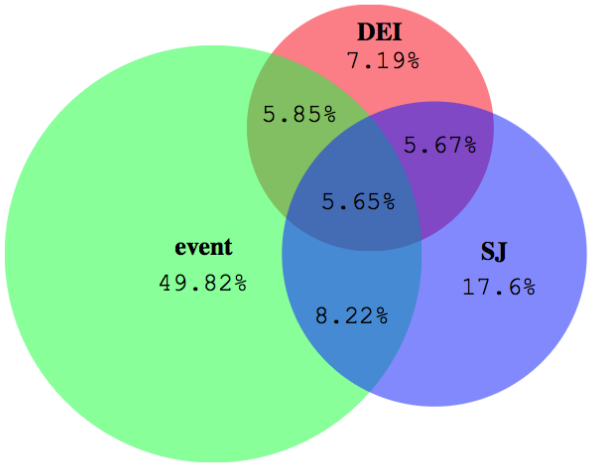


Figure 50. Comparision of DAS gens detected at different levels. Venn diagram created using Biovenn (Hulsen *et al*., 2008).

As mentioned in section 4.3.4.1.1 (page 151) when using DEI to identify potential DAS genes, if all isoforms in a gene are DE, this could be because the gene is undergoing DGE rather than DAS. Potentially the 7.19% of uniquely detected DAS genes by the isoform level analysis could be representing DEG. To investigate this, the percentage of genes where all isoforms of a specific gene model are differentially expressed in response to *B. cinerea* was determined for each area of the Venn diagram in Figure 50. Those genes detected to undergo DAS uniquely at the isoform level (17.6 % of genes), do not contain more gene models where all isoforms are differentially expressed compared to the other areas of the Venn diagram. This indicates that this area does indeed represent DEI and not DEG (Table 17).

Table 17. Comparison of percentage where DEI represents all isoforms in a gene by level. Column one gives the intersect being investigated, and column two depicts the percentage of genes where all known isoforms are differentially expressed in response to *B. cinerea* 24 hpi.

Set	Percentage of genes where all isoforms are DEI
DEI only	12.1%
DEI and event	7.2%
DEI and SJ	13.7%
All	10.0%

The SJ and the isoform level of analysis detected fewer of the validated transcripts compared to the event level analysis, potentially indicating that the event level

analysis is the most efficient at detecting *B. cinerea*-mediated DASed genes. However, the large number of uniquely detected *B. cinerea*-mediated DASed genes detected using the event level analysis could potentially be due to false positives. Comparing the top 100 and top 200 statistically significant *B. cinerea*-mediated DASed genes from each level of analysis with all statistically significant *B. cinerea*-mediated DASed genes found in the other two levels shows that a large proportion of the top DASed genes are uniquely detected by one method (Table 18).

Table 18. Percentage of top 100 and top 200 statistically significant genes uniquely detected at one level. Column one indicates the level of analysis, column two and three detail how many genes from the top 100 (column two) or top 200 (column three) DAS genes with the lowest p-value from that level of analysis are not detected to undergo DAS by the other two levels of analysis.

	Not detected in other two levels	
Level	Top 100	Top 200
DEI	40% (40/100)	41% (82/200)
Event	44% (44/100)	49.5% (99/200)
SJ	30% (30/100)	30.5% (61/200)

This indicates that for each level of analysis there are highly statistically significant genes uniquely detected by that level, adding weight to the theory that more than one level of analysis should be utilised to detect DAS.

To further investigate the differences between the levels of analysis the 924 genes detected to undergo *B. cinerea*-mediated DAS by Cooke (2013) were utilised. Of the 924 genes detected to undergo *B. cinerea* mediated DAS by Cooke (2013), 660 (71.4%) were also detected in the 7578 genes we detected to undergo *B. cinerea*-mediated DAS at 24 hpi. Of these 660 genes, 84.1% were detected at the event level with 50.9% being uniquely identified by the event level analysis. Both the isoform and SJ level of analysis uniquely identified a small proportion of the 660 genes, 3.9% and 8.2% respectively. This combined with the fact that the event level analysis detected four out of the five of the validated genes that undergo *B. cinerea*-mediated DAS indicates that the event level analysis is the most efficient at detecting DASed genes.

Of particular importance to note is that for the remainder of the analysis in this chapter, unless otherwise stated, genes detected to undergo DAS by all three levels of analysis will be utilised. This is because each level uniquely provides some highly statistically significant *B. cinerea*-mediated DAS genes. This DAS dataset includes the 1846 DAS genes detected by the DEI pipelines, Tophat2->CuffDiff2, STAR->CuffDiff2 and Sailfish->DESeq2, the 5269 DAS genes detected at the event level by the Sailfish->SUPPA with diffSplice pipeline, and the 2815 DAS genes from the intersect of the two splice junction pipelines, resulting in a list of 7578 *B. cinerea*-mediated DAS genes.

4.3.7 *B. cinerea* infection results in DAS of known defence-related genes in Arabidopsis

Of the 22302 Arabidopsis intron-containing genes, 34% undergo *B. cinerea*-mediated DAS (7578 out of 22302). This DAS dataset includes some well-known *B. cinerea* defence-related genes that cover a wide range of functions, including but not limited to, recognition of pathogens (*CERK1*, *LIK1*, *PBL27* and *BAK1*), regulation of gene expression (*MYB51*, *WRKY3*, *WRKY28*, *WRKY40* and WRKY60), and genes that encoding peptides with antimicrobial properties such as *PDF1.2*.

In order to obtain a functional profile of all the genes that undergo *B. cinerea*-mediated DAS, GO enrichment analysis was performed using BINGO 3.0.3 (Maere *et al*., 2005). GO terms enriched include multiple defence related GO terms, such as "defence response" and "response to biotic stimulus" including some fungus specific terms such as "response to fungus", "defence response to fungus" and "response to chitin", suggesting that that *B. cinerea*-mediated DAS genes are likely to be functionally significant in the Arabidopsis defence response. In addition, there is enrichment for GO terms related to RNA processing including splicing associated terms, potentially indicating that one of the underlying biological processes affected by *B. cinerea*-mediated DAS is the regulation of splicing.

Here we will analyse *B. cinerea*-mediated DAS in two defence related but distinctly different genes families, before beginning to elucidate potential regulatory mechanisms by investigating splicing factor associated genes and SRE motifs.

4.3.8 JAZ genes are DAS in response to *B. cinerea*

Moreno *et al* (2013) have shown that AS of *JAZ10* may be utilised to help regulate the negative feedback control of jasmonate signalling. This, combined with the fact that in the Cooke (2013) dataset there are three JAZ genes that are DASed in response to *B. cinerea* (*JAZ5*, *JAZ9* and *JAZ10*), indicates that this could be a good family of genes to investigate the effects of *B. cinerea*-mediated DAS.

There is a large range in transcript abundances of *JAZ* genes, with the majority of *JAZ* genes having more than one transcript detected. Some JAZ genes have large total transcript abundances in infected samples, such as *JAZ1*, *JAZ6* and *JAZ9*, whilst others are lowly expressed such as *JAZ4* and *JAZ13* (Figure 51). In the majority of JAZ genes one transcript accounts for the vast proportion of the total transcript abundances, with obvious exceptions being *JAZ3* and *JAZ10* (Figure 51).

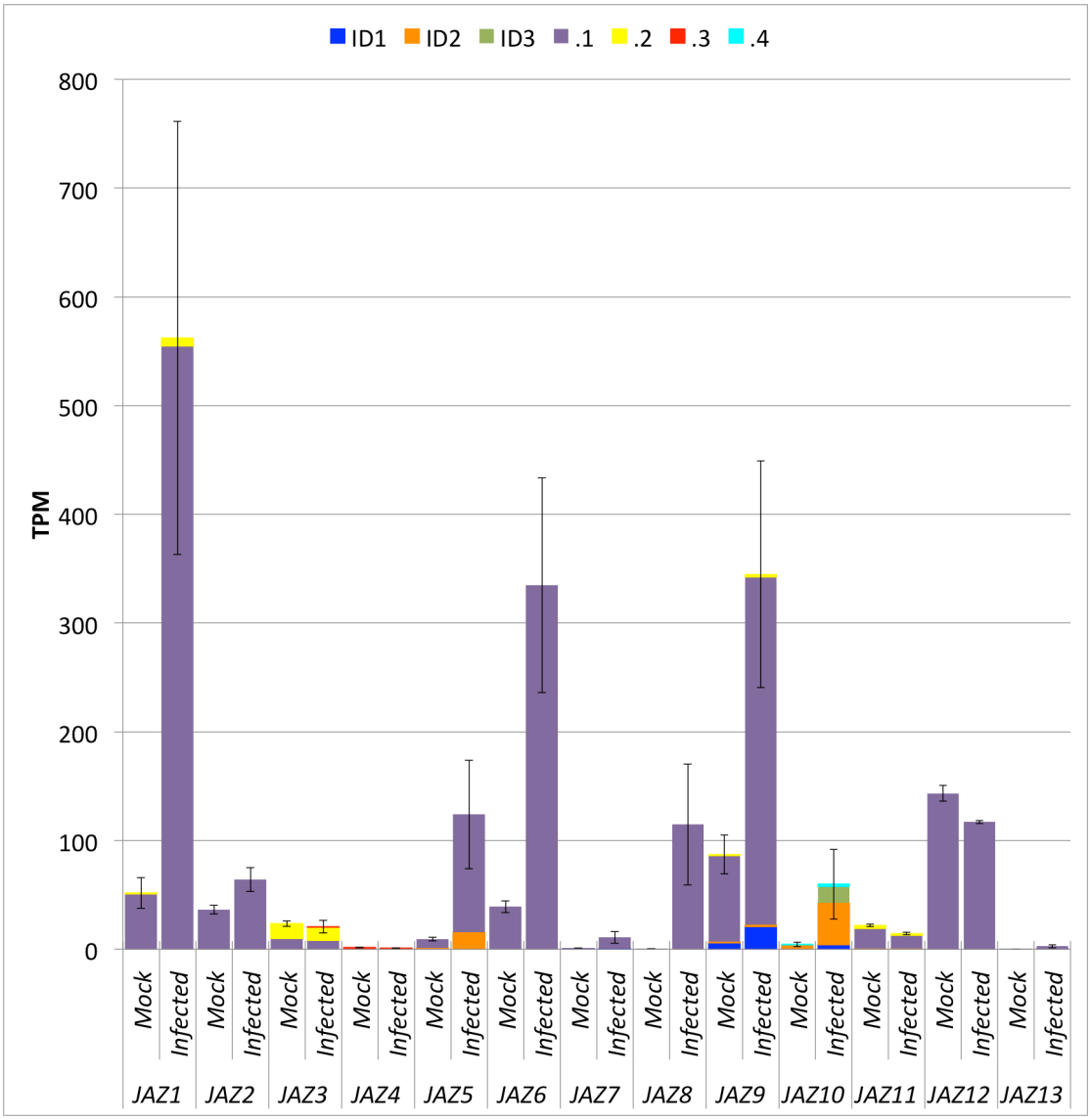


Figure 51. Total abundance of transcripts in each JAZ gene in mock and *B. cinerea* infected Arabidopsis leaves. *JAZ13* has the least amount of total transcript in mock samples with *JAZ12* having the largest. In *B. cinerea* infected samples *JAZ4* has the smallest total amount of total transcript with *JAZ1* having the largest. For the majority of JAZ genes multiple transcripts are detected in at least one condition (mock or *B. cinerea* 24 hpi). Error bars represent SEM of total transcript abundance from three biological replicates. Transcript abundances calculated as TPM utilising Sailfish.

4.3.8.1 JAZ genes can undergo Differential Alternative Splicing and Differential Gene Expression in response to B. cinerea infection

The majority (nine out of 13) of JAZ genes are DAS in response to *B. cinerea* 24 hpi, with six of these nine also undergoing differential gene expression (DGE) (Table 19).

Table 19. The majority of the *JAZ* genes are DASed in response to *B. cinerea* 24hpi. Column one provides the gene identification with column two detailing the gene name. Group (column three) refers to the subclade assigned by using phylogenetic analysis of the TIFY family (Bai *et al*., 2011). Column four identifies which JAZ isoforms are DAS and column five identifies which are DE. – indicates that the gene does not undergo DAS or DE in response to *B. cinerea* 24 hpi depending on column. *JAZ11* was determined to be DASed at the SJ level, the differential splice junction usage could have arisen from either *JAZ11.1* or *JAZ_ID2* as indicated by *

Gene ID	Name	Group	DAS	DEG
AT1G19180	JAZ1	I	JAZ1.1	YES
AT1G74950	JAZ2	I	JAZ2.1	-
AT3G17860	JAZ3	V	JAZ3.3	-
AT1G48500	JAZ4	V	-	-
AT1G17380	JAZ5	I	JAZ5.1 *JAZ5*_ID2	YES
AT1G72450	JAZ6	I	-	YES
AT2G34600	JAZ7	IV	JAZ7.1	YES
AT1G30135	JAZ8	IV	-	YES
AT1G70700	JAZ9	V	JAZ9.1	YES
AT5G13220	JAZ10	III	JAZ10_ID1 JAZ10_ID2 JAZ10_ID3	YES
AT3G43440	JAZ11	II	JAZ11.1* *JAZ12*_ID2*	-
AT5G20900	JAZ12	II	-	-
AT3G22275	JAZ13		JAZ13.1	YES

The eight genes that we detected to undergo DGE (*JAZ1*, *JAZ5*, *JAZ6*, *JAZ7*, *JAZ8*, *JAZ9*, *JAZ10*, and *JAZ13*) were also shown to be DE around 24 hpi in the *B. cinerea* time-series experiment carried out previously, with no additional *JAZ* genes detected to undergo DGE (Windram *et al*., 2012). This indicates that we have accurately detected which *JAZ* genes are DE. Out of the nine genes detected to undergo DAS in response to *B. cinerea* three of these were also detected to undergo DAS by Cooke (2013) (*JAZ5*, *JAZ9* and *JAZ10*), providing additional experimental evidence that these genes are DASed in response to *B. cinerea*. Only two genes are not detected to undergo DAS or DGE in response to *B. cinerea* 24 hpi (*JAZ4* and *JAZ12*) indicating that the JAZ genes are highly regulated in response to *B. cinerea* 24 hpi.

4.3.8.2 B. cinerea infection alters transcript ratios of JAZ genes

Most JAZ genes only have one transcript that is DASed in response to *B. cinerea* 24hpi (*JAZ1*, *JAZ2*, *JAZ3*, *JAZ7*, *JAZ9* and *JAZ13*) with only two genes showing statistically significant DAS in more than one transcript per JAZ gene (*JAZ5* and *JAZ10*). One DASed *JAZ* gene, *JAZ2*, only has one transcript detected in our analysis, potentially indicating that DAS of *JAZ2* could represent DGE. However, for the other eight DASed JAZ genes, more than one transcript was detected (in either condition). Changing the amount of one transcript can affect the relative ratio of transcripts that make up the gene expression level, potentially helping to fine-tune plant defence responses.

Of these eight genes, there is a negligible proportional change in isoform ratio of *JAZ1* (0.01%) and only very little change (1%) change in *JAZ5*. Although there are large increases in the abundances of both the *JAZ1.1* and *JAZ5.1* isoforms in response to *B. cinerea* 24 hpi the second isoform of each gene (*JAZ1.2* and *JAZ5_ID2*) also increased. However these genes are only expressed in relatively low amounts (Figure 52).

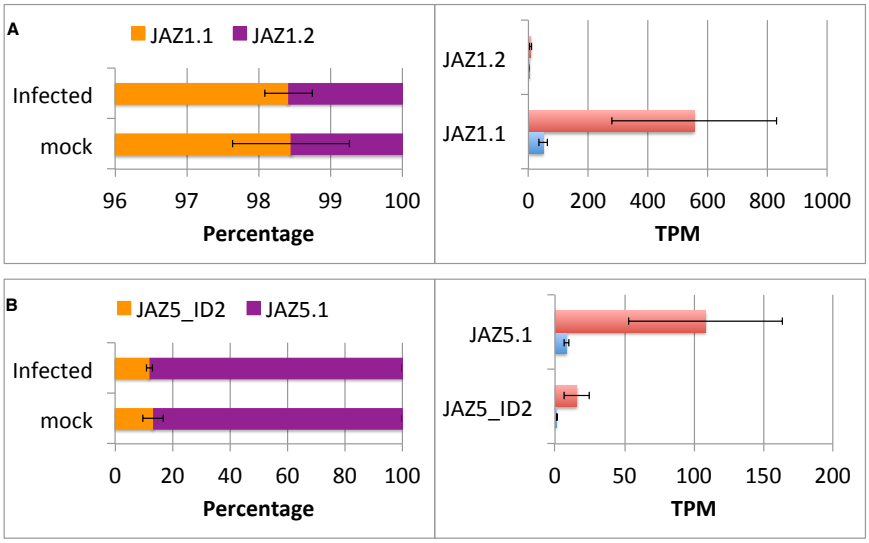


Figure 52. Proportional change in the ratios and total abundance of the *JAZ1* and *JAZ5* isoforms. There is a negligible proportional change in isoform ratios of *JAZ1* (0.01%) and only very little change (1%) change in *JAZ5* (left-hand panel). However there are large increases in the abundances of both the *JAZ1.1* and *JAZ5.1* isoforms, and the second isoform of each gene (*JAZ1.2* and *JAZ5_ID2*) also increase (right-hand panel). In the left-hand panel blue bars represent mock abundances and red bars represent infected. Error bars represent SEM of three biological samples.

There is only a very small amount of the transcript *JAZ1.2*. This combined with its relatively large variation between samples means that it was not determined to be a statistically significant DEI, resulting in *JAZ1* being classified as DAS in response to *B. cinerea* 24 hpi. The very small proportional change of transcripts indicates that the mechanism of regulation of this gene could be more concerned with gene expressional changes rather than DAS. Both *JAZ5* isoforms were detected to be DEI in response to *B. cinerea*, which could also indicate that this gene is predominantly regulated by DGE mechanisms. However additional evidence that *JAZ5* does indeed undergo *B. cinerea*-mediated DAS is that it was also detected to be DASed in the Cooke (2013) dataset. Thus, the 1% change in proportions could represent a DAS event. Further work on additional samples utilising additional techniques specific to investigating DAS of this gene could help to determine if *JAZ5* undergoes both DE and/or DAS in response to *B. cinerea*.

For three of the DASed *JAZ* genes, *JAZ7*, *JAZ9* and *JAZ10*, all of the isoforms detected show an increase in response to infection. However there is a larger proportional change of isoforms indicating that these genes may be subject to mechanisms that regulate both DAS and DGE. The proportion of *JAZ7_ID1* decreases in response to *B. cinerea* infection whilst *JAZ7.1* increases by 3.2%. Both isoforms increase in abundance in response to infection (Figure 53A). The proportion of *JAZ9.1* increases by 4.5% whilst the other proportions of the other isoforms decrease, all isoforms increase in abundance in response to infection ((Figure 53B). The proportion of *JAZ10_ID3* increases by 8%, whilst the other isoform decrease in proportion, *JAZ_ID1* has the smallest proportional change in isoform ratio (0.8%) two other isoforms

JAZ_ID2 and *JAZ10.4* also have a small proportional change (1.7% and 5.9% respectively) (Figure 53C).

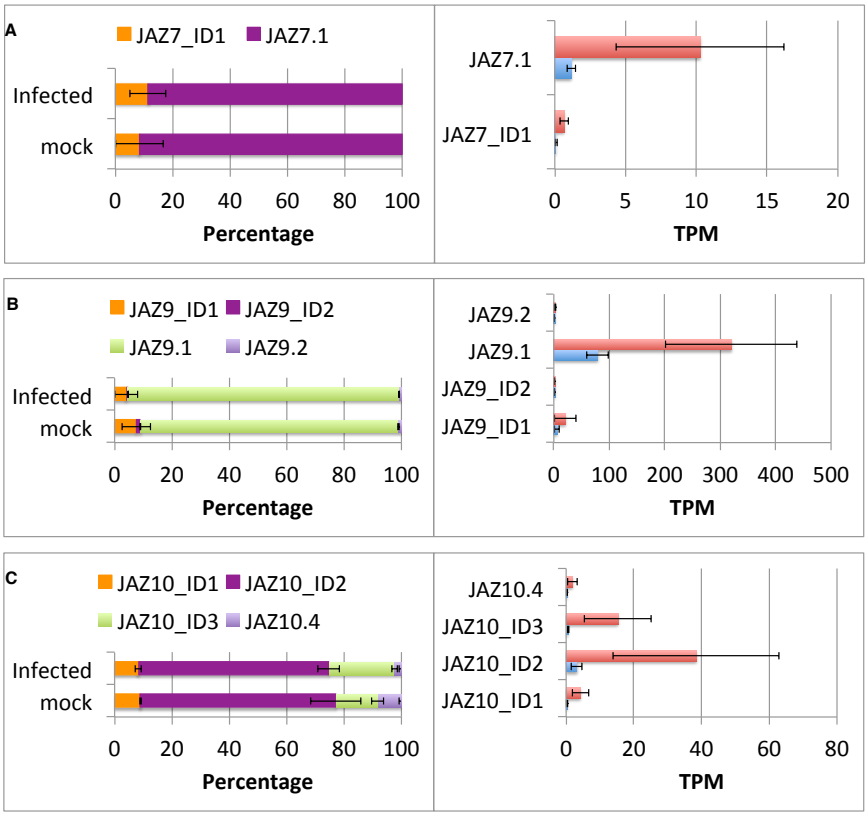


Figure 53. Proportional change in the ratios and total abundance of the *JAZ7*, *JAZ9* and *JAZ10* isoforms. Left hand panels show the proportion of each isoform. In response to *B. cinerea* infection there are proportional changes in the ratios of the *JAZ7* (A), *JAZ9* (B) and *JAZ10* (C) isoforms. The right-hand panels show the total abundance of each isoform, blue bars represent mock abundances, and red bars represent infected. Error bars represent SEM of three biological samples.

The idea that these genes may be subjected to mechanisms that regulated both DAS and DGE is supported by our analysis detecting both statistically DGE and DAS in response to *B. cinerea* 24 hpi. *JAZ7* and *JAZ10* were also detected to be DAS in the Cooke (2013) dataset supporting our DAS data, with Windram *et al* 2012 detecting DE of these three genes around 24 hpi supporting our DEG dataset.

JAZ3 and *JAZ11* were not detected to be DE in response to *B. cinerea* but do undergo DAS. Both genes have proportional changes of the transcripts in response to *B. cinerea* infection, with one transcript (*JAZ3.3*) only being detected in infected samples. There are only small changes in the transcript abundances with each gene

producing some transcripts that increase in response to *B. cinerea* infection whilst others decrease (Figure 54 right hand panels). A. *JAZ3* has one isoform that is only seen in mock samples (*JAZ3_ID2*) (although it is a very small percentage of the total transcripts (0.08%)), and one transcript that is only detected in infected sample (*JAZ3.3*), this also represents a relatively small percentage of the total transcripts (2.8%) (Figure 54A). *JAZ3.1* does not significantly alter between treatments whereas there is a small proportional change in *JAZ3.2* (3.8% decrease in **B. cinerea** infected samples) (Figure 54A, left-hand panel). The proportional change of *JAZ11* transcripts is also quite small with a decrease of 3.3% of *JAZ11_ID2* accompanied by an increase of 2.1% and 1.2% of *JAZ11.1* and *JAZ11.2* respectively (Figure 54B, left-hand panel).

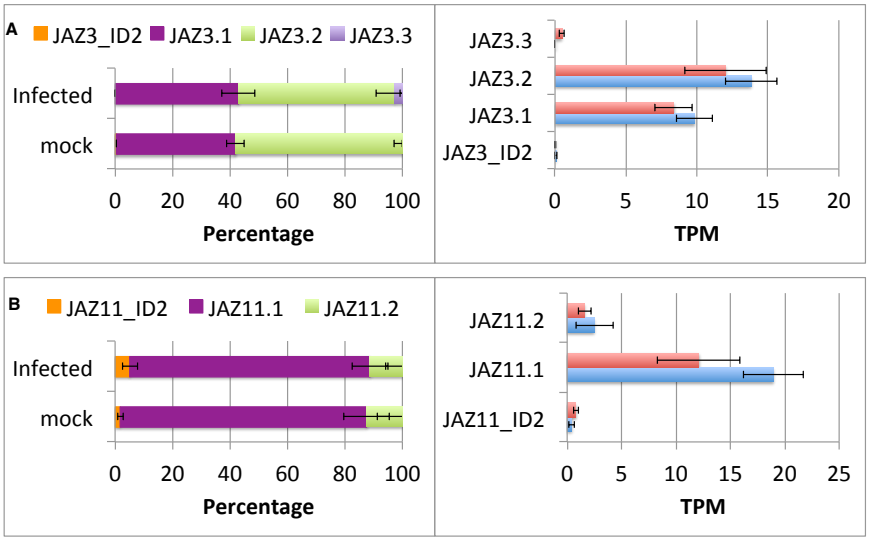


Figure 54. Proportional change in the ratios and total abundance of the JAZ3 and JAZ11 isoforms in response to *B. cinerea* 24 hpi. In response to *B. cinerea* 24 hpi there is a proportional change in isoforms originating from both *JAZ3* (A) and *JAZ11* (B) (left-hand panels). This is accompanied by some transcripts increasing in abundances whilst other decrease (right hand panels). In the right-hand panel blue bars represent mock abundances and red bars represent infected. Error bars represent SEM of three biological samples.

This could indicate that these JAZ genes are predominantly regulated by DAS mechanisms, but the overall changes are very small and thus this could represent noise rather than a biological phenomena. This could also be the case for *JAZ13*, because only very small overall amounts were detected (<3 TPM). Further experiments targeting these genes would be required to confirm if they are undergoing *B. cinerea*-mediated DAS.

4.3.8.3 In response to B. cinerea infection JAZ3 isoforms with shorter 5'UTRs are utilized.

The TIFY and ZIM domains of the three main splice variants of *JAZ3* are all the same, with the differences occurring in the 5'UTR and the first exon. JAZ3.1 contains a large intron in the first exon and has the shortest 5'UTR. *JAZ3.2* and *JAZ3.3* may represent incomplete transcripts, with the RI events potentially indicating that they have multiple uORFs. This in turn could target them for NMD. The small intron in *JAZ3.3* appears to alter the translational start site (Figure 55).

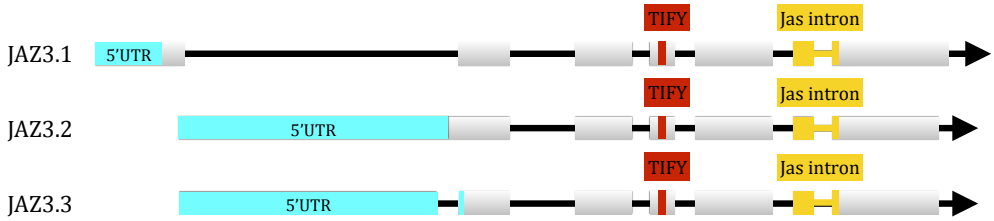


Figure 55. Gene models for *JAZ3*. Grey boxes represent exons, black lines represent introns, red boxes represent TIFY motif and yellow boxes represent Jas domains. A yellow line represents the Jas intron.

The length of an intron in the 5'UTR enhances gene expression in a size dependant manner, the longer the intron the greater the gene expression (B. Y. W. Chung *et al*., 2006). *JAZ3.1* does not contain an intron in the 5'UTR and has the shortest 5'UTR, *JAZ3.2* also does not contain an intron in the 5'UTR and the 5'UTR is longer than *JAZ3.1*, *JAZ3.3* has the longest 5'UTR but it does contain an intron, although even with the intron spliced out its 5'UTR is the longest. *JAZ3.3* was detected to undergo DAS in response to *B. cinerea*, with increased expression of this transcript in infected samples. This could potentially be a way of reducing the gene expression of the *JAZ3* in response to *B. cinerea*, resulting in greater transcription of JA dependent defence-related genes, if the length of the intron in the 5'UTR affects gene expression in a size-dependant manner. In addition, the RI event in *JAZ3.2* and *JAZ3.3* results in multiple AUG-initiated uORFs. This could potentially target these transcripts for NMD, which could also affect gene expression levels. This gene is not DE in response to *B. cinerea* 24 hpi, potentially indicating that expression levels of *JAZ3* are regulated at the transcript level.

4.3.8.4 DAS of JAZ genes has the potential to fine-tune the negative feedback control of jasmonate signalling in response to B. cinerea infection.

It has previously been shown that different isoforms of *JAZ10* can affect desensitization to JA in Arabidopsis. Moreno *et al* (2013) showed that *JAZ10.4* lacks the Jas motif and is thought to be involved in a negative feedback control of jasmonate signalling. *JAZ10* has four isoforms detected in our analysis, *JAZ10_ID1*, *JAZ10_ID2*, *JAZ10_ID3* (which replaces *JAZ10.1*, *JAZ10.2* and *JAZ10.3* in TAIR10, they have the same exon/intron combinations but minor amino acid changes), and *JAZ10.4*. The fully spliced isoform, *JAZ10.2*, contains a TIFY motif, and a JAS domain that is divided in two by the fourth intron. This intron is known as the Jas intron (H. S. Chung *et al*., 2010). *JAZ10_ID1* and *JAZ10_ID3* have all or part of the Jas intron retained (by a RI and A5SS event respectively) resulting in a PTC at the beginning of this intron, which is predicted to produce a protein with a truncated Jas domain. *JAZ10.4* has an A5SS of intron 3 causing a frame shift that results in proteins with no Jas domains. (Figure 56).

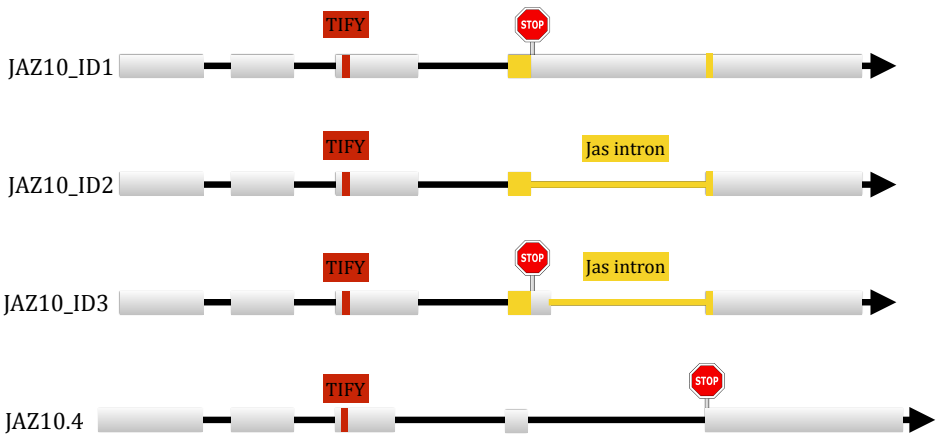


Figure 56. *JAZ10* gene models. Grey boxes represent exons, black lines represent introns, red boxes represent TIFY motif and yellow boxes represent Jas domains. A yellow line represents the Jas intron.

In response to *B. cinerea* 24 hpi the main changes in proportion of the isoforms is seen between *JAZ10.4*, which lacks the Jas motif and decreases, and *JAZ10_ID3*, which has a truncated Jas domain and increases. The proportion of the fully spliced isoform, *JAZ10_ID2*, only increases slightly (1.9%). However the combined isoform expression increase approximately four fold (log_2), resulting in a much larger amount

of the fully spliced transcript (Figure 53C). JAZ isoforms that lack the Jas motif cannot interact with COI1. However, those that have a truncated Jas motif have been shown to be able to interact with COI1, but the interaction is significantly decreased compared to those isoforms with a fully intact Jas domain (H. S. Chung *et al*., 2010). One potential mechanism of how DAS of *JAZ10* can fine-tune the negative feedback control could be that in response to *B. cinerea* infection, the isoforms that cause negative feedback i.e. those which cannot interact with COI1 (*JAZ10.4*) are reduced. However to prevent overstimulation of the pathway some negative feedback is still required. Thus the *JAZ10_ID3* isoform that has reduced ability to interact with COI1 is utilised, reducing the amount of negative feedback compared to when utilising JAZ10.4, but not abolishing it completely.

JAZ5 also has the potential to utilise DAS to reduce the negative feedback in the jasmonate signalling pathway. In the RNA-Seq experiment two isoforms were detected to be DE, *JAZ5.1* and *JAZ5_ID2*. *JAZ_ID2* retains the Jas intron causing a PTC, which we predict will cause a protein with a truncated Jas domain (Figure 57). In response to *B. cinerea*, both isoforms are increased. However, there is a small proportional change in the isoform ratios. In infected samples, 1% less of the total transcripts is comprised of *JAZ5_ID2* compared to mock samples. Due to the truncated Jas domain we predict that proteins encoded by *JAZ5_ID2* modulate the negative feedback in a similar manner as demonstrated to occur in *JAZ10_ID3*. By reducing the proportion of *JAZ5_ID2* in response to *B. cinerea*, the negative feedback is reduced.

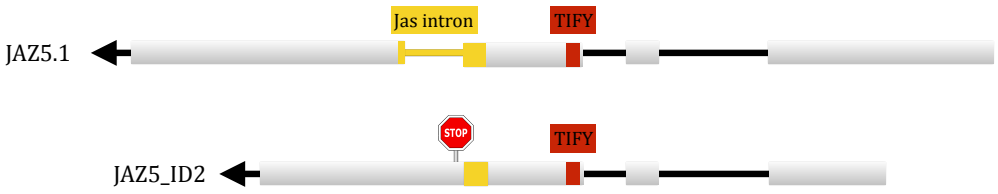


Figure 57. *JAZ5* gene models. Two isoforms of *JAZ5* were detected in our RNA-Seq experiment; *JAZ5.1* and *JAZ5_ID2*. *JAZ5_ID2* retains the Jas intron resulting in a PTC at the beginning of this intron, possibly producing a protein with a truncated Jas domain. Grey boxes represent exons, black lines represent introns, red boxes represent TIFY motif and yellow boxes represent Jas domains. A yellow line represents the Jas intron.

4.3.8.5 *Premature termination codons coupled to nonsense mediated decay in JAZ genes is a potential regulatory mechanism of the jasmonate signalling pathway during the defence response*

Although a role for the NMD pathway regulating JAZ mRNA accumulation has to our knowledge not previously been identified, two JAZ encoding genes, *JAZ9* and *JAZ11*, have AS gene models that are likely to be targeted for NMD. *JAZ9* has four isoforms (*JAZ9.1*, *JAZ9.2*, *JAZ9_ID1* and *JAZ9_ID2*). Compared to *JAZ9.1*, *JAZ9.2* has an additional intron in the fifth exon, *JAZ9_ID1* skips exons three and four, and *JAZ9_ID2* has an A3′ SS in exon 4, and an A5SS in exon 5. The isoform *JAZ9.1* and *JAZ9.2* have intact TIFY and Jas motifs, with *JAZ9.2* having 24 fewer amino acids between the two domains due to the additional intron in the fifth exon. The alternative splicing events in *JAZ9_ID1* and *JAZ9_ID2* result in PTCs (Figure 58).

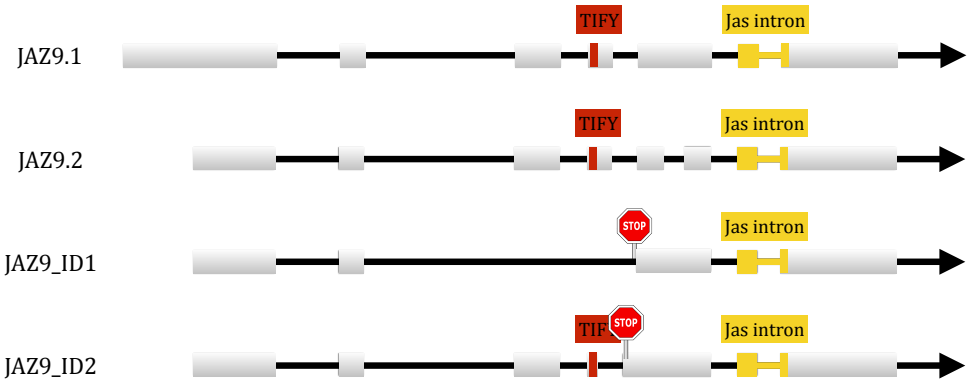


Figure 58. JAZ9 gene models. Four isoforms of *JAZ9* were detected in our RNA-Seq experiment; *JAZ9.1*, *JAZ9.2*, *JAZ9_ID1* and *JAZ9_ID2*. *JAZ9.1* and *JAZ9.2* encode proteins with intact TIFY and Jas domains, the difference between the isoforms being 25 amino acids between the two domains. *JAZ9_ID1* and *JAZ9_ID2* contain a PTC before the Jas domain due to a skipped exon (*JAZ_ID1*) and A3SS (*JAZ_ID2*). Grey boxes represent exons, black lines represent introns, red boxes represent the TIFY motif and yellow boxes represent Jas domains. A yellow line represents the Jas intron.

In both mock and infected samples, the majority of the isoforms are *JAZ9.1,* and in response to infection the proportion of this isoform increased whilst those isoforms that contain PTCs decrease (*JAZ9_ID1* and *JAZ9_ID2*). The isoforms that contain the PTC are likely to undergo NMD because they have large 3'UTR (>350nt) and

downstream splice junctions (Kalyna *et al*., 2012). Thus, in response to infection those isoforms producing functional transcripts are increased. In *JAZ11* the opposite occurs, functional isoforms decrease, whilst those likely to be targeted by the NMD pathway increase.

JAZ11 has three isoforms in the RtATD dataset, *JAZ11.1, JAZ11.2* and *JAZ11_ID2*. *JAZ11.1* and *JAZ_ID2* have eight exons and seven introns with *JAZ_ID2* having an A3SS in the second intron. This A3SS results in a PTC that is likely to be coupled to NMD due to the long 3'UTR that it creates (>350nt) and downstream splice junctions (Kalyna *et al*., 2012). The *JAZ11.1* isoform is predicted to encode a protein with two TIFY and two Jas domain motifs, whereas *JAZ11.2*, which skips exons 3, 4 and 5, only encodes one TIFY and Jas domain motif (Figure 59).

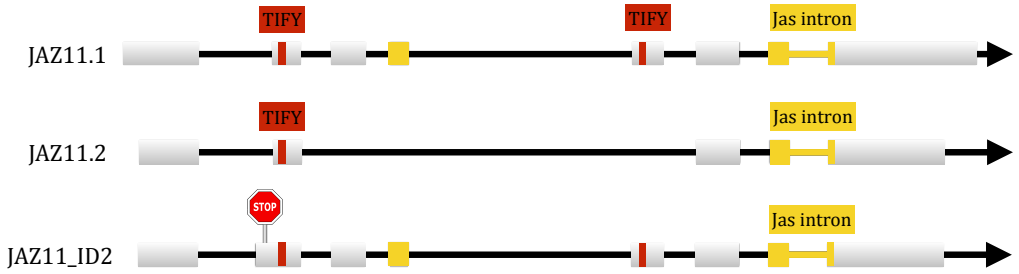


Figure 59. *JAZ11* gene models. Three isoforms of *JAZ11* were detected in our RNA-Seq experiment: *JAZ11.1*, *JAZ11.2* and *JAZ11_ID2*. *JAZ11.1* has two TIFY and two Jas domains, whilst *JAZ11.2* only has one of each due to skipped exons. *JAZ11_ID2* contains a PTC before the TIFY domain due to an A3SS. Grey boxes represent exons, black lines represent introns, red boxes represent the TIFY motif and yellow boxes represent Jas domains. A yellow line represents the Jas intron.

In response to *B. cinerea* 24hpi two splice junctions (15369198-15369296 and 15368087-15368183) show statistically significant decreases in usage, although the fold change is relatively low (0.8 $\log_2$ fold change). These splice junctions do not occur in *JAZ11.2* indicating that in response to *B. cinerea* infection either one or both of *JAZ11.1* or *JAZ11_ID2* is reduced in usage. Investigating the isoform ratios showed that there is an increase in the proportion *JAZ11_ID2* in response to infection, accompanied by a decrease in the proportions of the functional transcripts (*JAZ11.1* and *JAZ11.2*). Further work would be needed to confirm this by designing experiments that can easily differentiate between *JAZ11.1* and *JAZ11_ID2*. Interestingly, *JAZ11* is not differentially expressed at the gene level; this combined

with the above could potentially indicate that the mechanism of regulation of JAZ11 is NMD coupled to DAS.

Investigating the JAZ gene family highlights that the potential effects of DAS covering two of the three known main consequences of AS, alterations in functional domains (truncated jas domains) and altering concentrations of proteins (PTC coupled to NMD and altered 5'UTR). To help elucidate if these are general effects of DAS or specific to this gene family, the LRR-RLK family is investigated.

4.3.9 Members of the Receptor Like Kinase family undergo *B. cinerea*-mediated DAS.

The RLKs are a large gene family with at least 610 members, of which 233 are LRR-RLKs (Shiu *et al*., 2004). Out of these 610 genes that encode RLKs we have found 183 to be DAS in response to *B. cinerea* 24 hpi. This is significantly more than the 18 that we found by mining the Cooke (2013) dataset in section 3.4.4 page 119. Of these 18, 16 are also found to be DAS in response to *B. cinerea* 24 hpi in our dataset. On average, total transcript abundances of RLK genes found to be DASed uniquely in our data were lower compared to those found in both this dataset and the Cooke (2013) dataset. This indicates that we are detecting DAS of transcripts with lower abundances, potentially indicating that our dataset is more sensitive than Cooke (2013).

GO enrichment analysis shows that these 183 *B. cinerea*-mediated DASed genes are enriched for functions related to four main categories; biological regulation, defence responses, localisation and transport, and metabolic processes. Expected defence response associated GO terms are enriched, such as "response to biotic stimulus" and "regulation of programmed cell death", including fungal specific terms such as "response to chitin" (Figure 60).

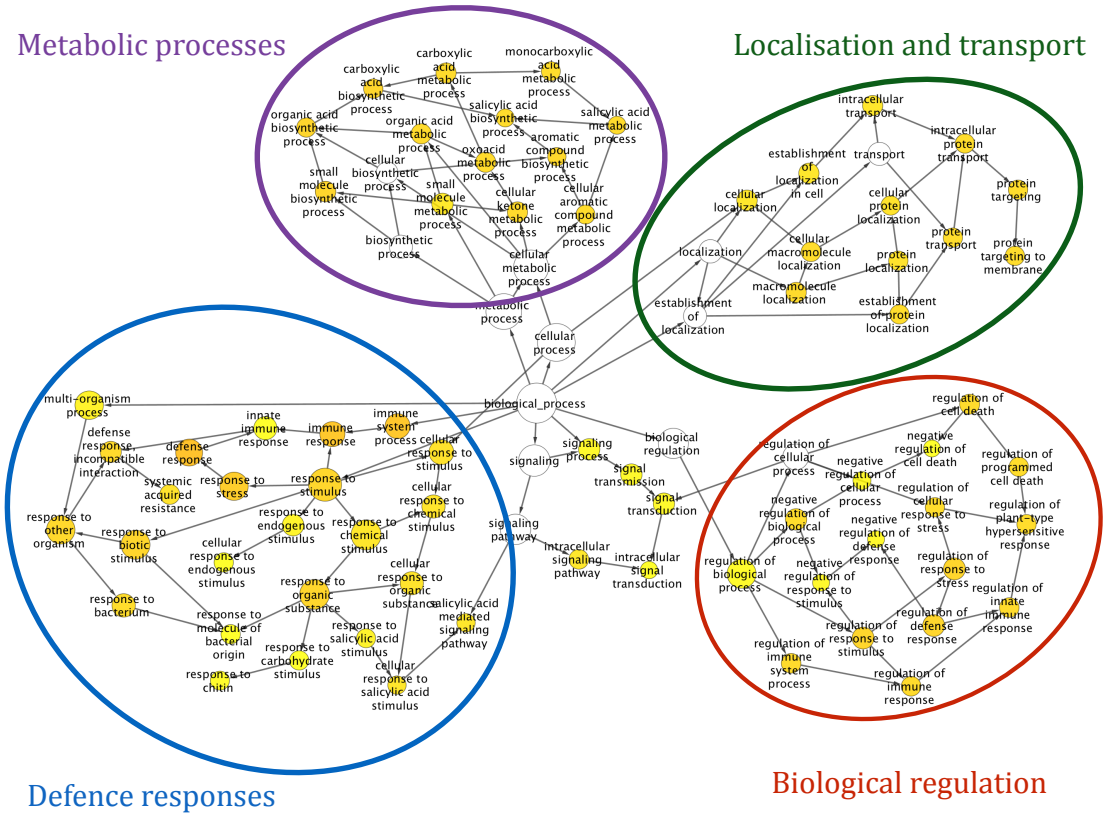


Figure 60. Network of biological processes of Gene Ontology (GO) categories for *B. cinerea*-mediated DAS Receptor Like Kinase genes. Figure generated by utilising Cytoscape 3.2.1 and BINGO 3.0.3 (Maere *et al.*, 2005; Shannon *et al.*, 2003).

In addition to RLK genes that undergo *B. cinerea*-mediated DAS being enriched for defence related GO terms, known defence related RLKs are DAS in response to *B. cinerea* 24hpi. This indicates that DAS may play a regulatory role in plant defence to necrotrophic pathogens.

4.3.9.1 CERK1 undergoes DAS in response to B. cinerea infection

The classic fungal PAMP chitin is detected by a variety of receptors in Arabidopsis. One of the most intensely studied is *CERK1* (*AT3G21630*) which has four isoforms in the AtRTD reference transcript dataset, *AT3G21630.1* (also in TAIR10), *AT3G21630_ID1*, *AT3G21630_ID3* and *AT3G21630_ID4*. *CERK1_ID1*, *CERK1_ID3* and *CERK1_ID4* have PTCs due to an A5SS in the third exon (*CERK1_ID1* and *CERK1_ID3*) and SEs in *CERK1_ID4* (Figure 61). This results in the transcripts having long 3'UTRs, which we predict will make them targets for the NMD pathway.

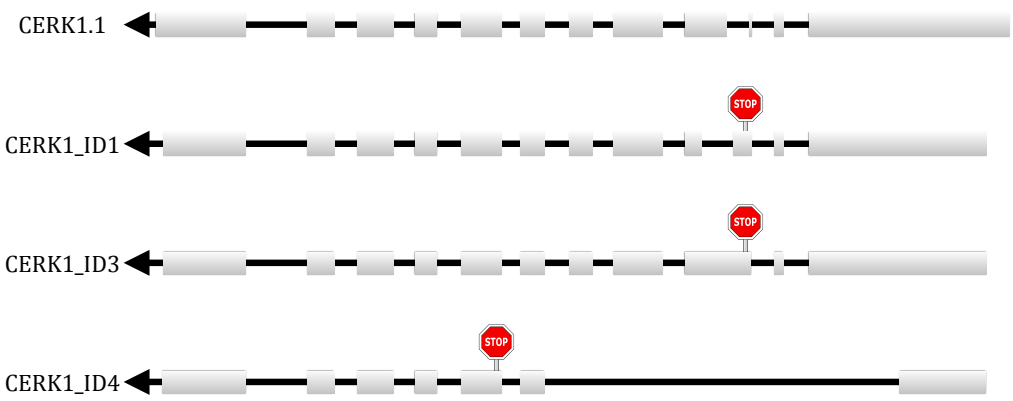


Figure 61. *CERK1* gene models. There are four isoforms of *CERK1* in the AtRTD reference transcript dataset *CERK1.1*, *CERK1_ID1*, *CERK1_ID3* and *CERK1_ID4*. *CERK1_ID1*, *CERK1_ID3* and *CERK1_ID4* contain a PTC caused by A5SS (*CERK1_ID1* and *CERK1_ID3*) and skipped exon (*CERK1_ID4*) events respectively. Grey boxes represent exons and black lines represent introns.

In our RNA-Seq experiment *CERK1* undergoes DAS in response to *B. cinerea* 24 hpi; with *CERK1.1* being the isoform detected to undergo DAS in response to *B. cinerea*. *CERK1_ID4* is not detected in either condition and *CERK1_ID3* is only present in very small quantities; *CERK1_ID3*, *CERK1_ID4* and *CERK1.1* all increase in abundance in response to *B. cinerea* 24 hpi (Figure 62A). Investigating the transcript ratios shows that in response to infection the proportion of the transcript that encodes the functional protein (*CERK1.1*) increases, whilst *CERK1_ID4*, that we predicted to be targeted for degradation by NMD, decreases (Figure 62B). This indicates that DAS coupled to NMD via creation of a PTC could be a regulatory mechanism of the RLK families.

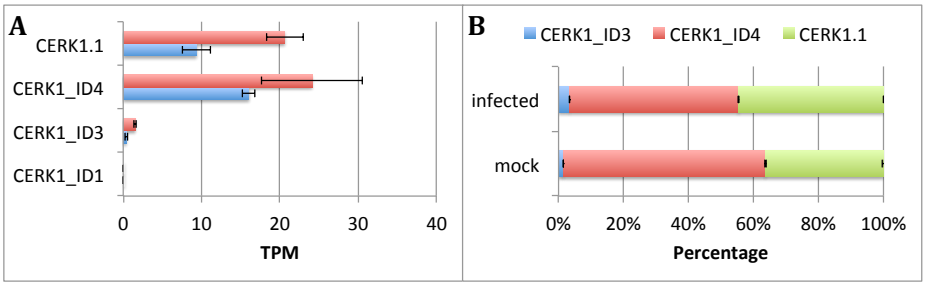


Figure 62. *CERK1* increases in abundance and has altered isoform ratios in response to *B. cinerea* 24 hpi. A. In response to infection *CERK1.1*, *CERK1_ID4* and *CERK1_ID3* increase in abundance. *CERK1_ID1* was not detected in either condition while *CERK1_ID3* is only detected in small amounts. B. In response to *B. cinerea* 24 hpi the proportion of *CERK1_ID4* decreases and *CERK1.1* increases. Blue bars represent mock abundances and red bars represent infected. Error bars represent SEM of three biological samples.

The RLK *PBL27* that is immediately downstream of *CERK1* and has been shown to play a role in the regulation of chitin-induced immunity (Shinya *et al*., 2014), also undergoes *B. cinerea*-mediated DAS. *PBL27* is DAS in response to *B. cinerea*, but it is not differentially expressed. Only one transcript (*PBL27.2*) increases its expression in response to *B. cinerea* 24 hpi resulting in a 23.4% increase in this transcript in infected samples (Figure 63).

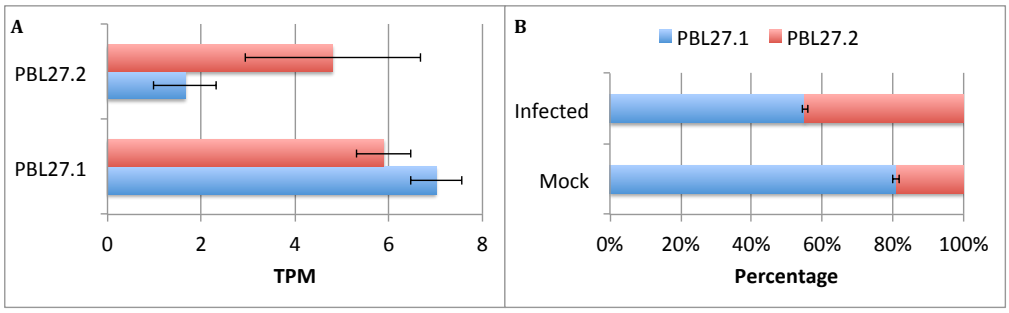


Figure 63. The proportion of *PBL27.2* increases in response to *B. cinerea* 24 hpi. A. In response to infection *PBL27.2* increased in abundance, whilst *PBL27.1* has no significant change (p=0.12 and p=0.090 Student t-test respectively). Blue bars represent mock abundances and red bars represent infected. B. In response to *B. cinerea* 24 hpi the proportion of *PBL27.2* increases by 23.4%. Blue and red proportions of bars represent *PBL27.1* and *PBL27.2* respectively. A and B. Error bars represent SEM of three biological samples.

The coding regions of *PBL27.1* and *PBL27.2* are the same. However, there is a difference in the length of the 5'UTR. *PBL27.1* contains an intron in this region that is retained in the AS isoform *PBL27.2* resulting in a longer 5'UTR in the *PBL27.2* isoform (Figure 64). This RI event contains a GUG-initiated uORF, which could potentially result in this transcript being targeted for NMD.

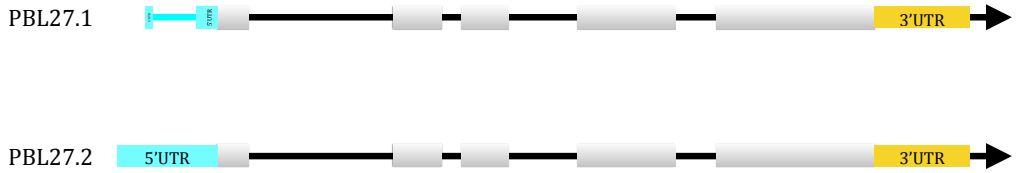


Figure 64. Transcript models for *PBL27*. Two *PBL27* isoforms were detected in the RNA-Seq experiment; *PBL27.1* and *PBL27.2*. The isoforms differ in their 5'UTR UTR regions. Grey boxes represent exons, black lines represent introns, cyan boxes and lines represent 5'UTR exons and introns respectively, with yellow boxes represent 3'UTR.

In response to *B. cinerea* infection the proportion of the *PBL27* isoform with the longest 5'UTR (*PBL27.2*) increased. The length of a genes 5'UTR is linked the level of gene expression, the longer 5'UTRs the higher the gene expression (B. Y. W. Chung *et al*., 2006). *PBL27* is not DE at the gene level, thus this is another potential example of DAS regulating gene expression levels of defence related proteins. AS coupled to NMD is another factor that could potentially be involved in increasing the gene expression levels for this gene.

4.3.9.2 LRR-RLKs are DAS in response to B. cinerea

The LRR-RLK subfamily consists of 233 members (Shiu *et al*., 2004). Of these LRR-RLKs genes, we detected 30% (71genes) to be DAS in response to *B. cinerea* 24hpi. Functional analysis of these DAS genes indicates that enriched GO terms fall into three main categories; biological regulation, defence responses, and localisation and transport. Enrichment of expected defence related GO terms such as "response to biotic stimulus" and "regulation of programmed cell death" are seen (Figure 65).

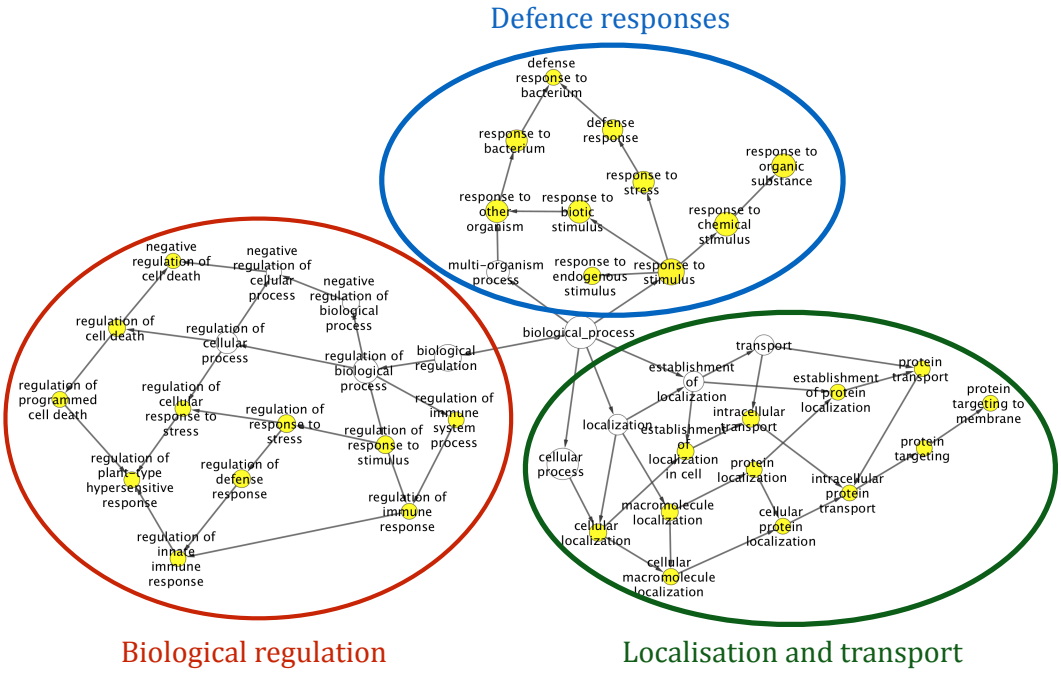


Figure 65. Network of biological processes of Gene Ontology (GO) categories for *B. cinerea*-mediated DAS LRR-RLK. Figure generated by utilising Cytoscape 3.2.1 and BINGO 3.0.3 (Maere *et al*., 2005; Shannon *et al*., 2003).

4.3.9.2.1 LRR-RLKs with known biological functions are DAS in response to *B. cinerea*

Gou *et al* (2010) identified 35 LRR-RLKs with known biological functions that are alternatively spliced in Arabidopsis leaf tissue. We identified 15 of these to undergo DAS in response to *B. cinerea* 24 hpi, including some well-known defence-related genes such as the BAK1 (*AtSERK3)*, *FLS2* and *ERF* genes (Table 20).

Table 20. Fifteen LRR-RLKs with known biological functions are DASed in response to *B. cinerea* 24hpi. Column one gives the gene identifier with the gene name in column two, LRR subfamily (column) three and function (column) four are taken from Shiu *et al* 2004 and Gou *et al* 2010 respectively (Gou *et al*., 2010; Shiu *et al*., 2004).

Gene Id	Name	Subfamily	Functions
AT4G29990	LRRPK	LRR I	Light signal transduction
AT1G34210	AtSERK2	LRR II	Male Sporogenesis
AT1G71830	AtSERK1	LRR II	BR signalling/ Male Sporogenesis
AT2G13790	BKK1/ AtSERK4	LRR II	BR signalling/Pathogen response/Cell death
AT4G33430	BAK1/ AtSERK3	LRR II	BR signalling/Pathogen response/Cell death

AT1G11130	Scrambled/SRF9/SUB/ Strubbelig	LRR V	Root epidermis patterning/Organ development/Positional signalling /cell morphogenesis
AT3G13380	BRL3	LRR X	Brassinosteroid receptor/Vascular differentiation
AT4G39400	BRI1	LRR X	Brassinosteroid receptor
AT5G07280	EMS1/EXS	LRR X	Anther development
AT5G48380	BIR1	LRR X	Cell death and innate immunity
AT1G75820	CLV1	LRR XI	Meristem differentiation and maintenance
AT4G28490	HAESA	LRR XI	floral organ abscission
AT5G20480	EFR	LRR XII	Pathogen response
AT5G46330	FLS2	LRR XII	Pathogen response
AT2G35620	FEI2	LRR XIII	Cell Wall Biosynthesis

4.3.9.2.2 Defence related LRR-RLKs are DAS in response to *B. cinerea*

The 71 *B. cinerea*-mediated DAS LRR-RLK genes contain some well-known defence related RLKs. These include the genes that encode the PRR, *EFR* and *FLS2*, one of the two functionally redundant receptors that perceive the DAMP *AtPEP1*, *PEPR1*, plus the plant defence response modulating gene *BAK1*/*SERK3* that has been shown to interact with all of these receptors, and is proposed to work as a modulator of the plant defence response (Greeff *et al*., 2012; Roux *et al*., 2011). Investigating how DAS can affect some of these well-known defence related genes could help to broaden our understanding of plant defence mechanisms.

4.3.9.2.2.1 DAS of BAK1 is likely to be coupled to the nonsense mediated decay pathway

In response to *B. cinerea* 24 hpi *BAK1.1* (*AT4G33430.1*) and *BAK1_ID3* are DASed. Although *BAK1* has five isoforms in the AtRTD reference transcript dataset, (*BAK1.1*, *BAK1.2*, *BAK1_ID1*, *BAK1_ID3* and *BAK1_ID4*), *BAK_ID4* is not detected and only negligible amounts of *BAK1.2* and *BAK1_ID1* are detected. In response to *B. cinerea* 24 hpi the abundance of *BAK1.1* significantly increases (p=0.017 Students t-test) whilst *BAK1_ID3* isoform decreases (p=0.1 Students t-test). These changes in transcript abundances results in an 18% increase in the proportion of *BAK1.1* in *B. cinerea* infected samples (Figure 66).

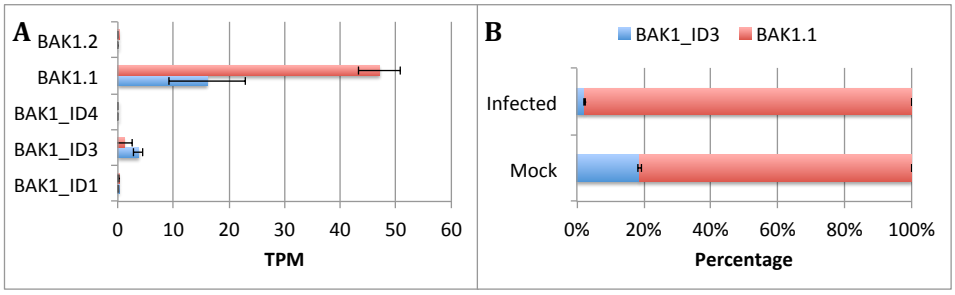


Figure 66. The proportion of *BAK1.1* increases in response to *B. cinerea* 24 hpi. A. Transcript abundancies in TPM. Blue and red bars represent mock and infected samples. B. Ratio of isoforms in mock and *B. cinerea* infected Arabidopsis leaves. A and B. Error bars represent SEM of three biological samples.

An A5SS event in the second exon of *BAK1_ID3* causes this exon to be 4 nucleotides shorter than the second exon in *BAK1.1*, resulting in a frame-shift that causes a PTC to occur in the middle of the third exon that we predict will target it for NMD (Figure 67).

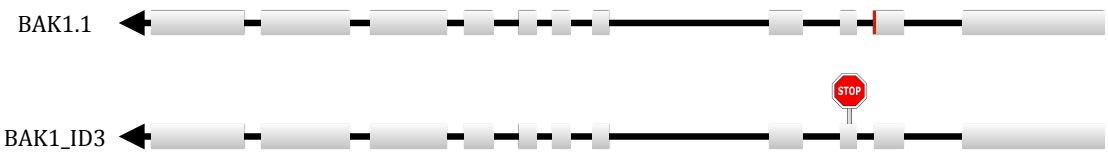


Figure 67. Gene models for *BAK1*. Two *BAK1* isoforms were detected in the RNA-Seq experiment: *BAK1.1* and *BAK1_ID3*. An A5SS of the second exon is predicted to result in a PTC in the third exon potentially targeting the isoform for NMD. Grey boxes represent exons, black lines represent introns, and the red area of BAK1.1 highlights the additional 4 amino acids present due to an A5SS event.

4.3.9.2.2.2 DAS of Leucine Rich Repeat Receptor Like Kinases can affect the tandem leucine rich repeats.

In response to *B. cinerea* 24 hpi, there is DAS of *ERF* in conjunction with DE. Two *ERF* isoforms were detected in our RNA-Seq experiment *ERF.1* (*AT5G20480.1*) and the AS transcript *ERF_ID1* which has an A5SS event. The A5SS event is predicted to cause the resulting protein to not encode the tandem leucine rich repeats domains; proteins from both isoforms are predicted to contain the Leucine Rich Repeat N-terminus domain (a cysteine rich domain that often flanks tandem leucine rich repeats) and the protein kinase catalytic domain (Figure 68).

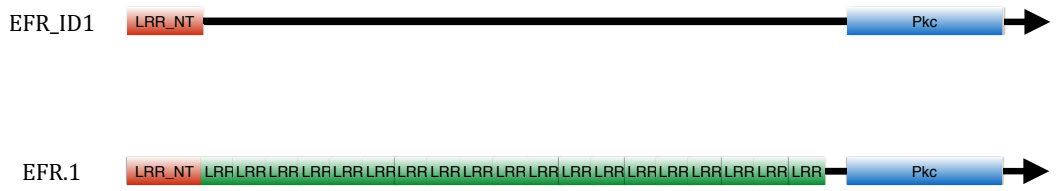


Figure 68. Gene models for *EFR* with predicted domain superimposed. Two *EFR* isoforms were detected in the RNA-Seq experiment; *EFR.1* and *EFR_ID1*. An A5SS of the first exon is predicted to result in the LRR tandem repeats to not be present in the protein encoded by EFR_ID1. Boxes represent exons, black lines represent introns, and the red area represents the LRR N terminal domain, the green areas represent LRR and the blue areas represent the protein kinase domain. Protein domains predicted using the Conserved Domain Database (Marchler-Bauer *et al*., 2015).

In response to *B. cinerea* 24 hpi, the abundance of both isoforms significantly increases (p=0.08 and 0.05, Students t-test for *ERF_ID1* and *ERF.1* respectively) (Figure 69A). However, there is a higher $\log_2$ fold change of *ERF.1* (*ERF_ID1* has an increase of 1 $\log_2$ fold change compared to *ERF.1* that has a 1.4 $\log_2$ fold change). This results in a nine per cent increase in the proportion of *ERF.1* in response to *B. cinerea* 24 hpi (Figure 69B).

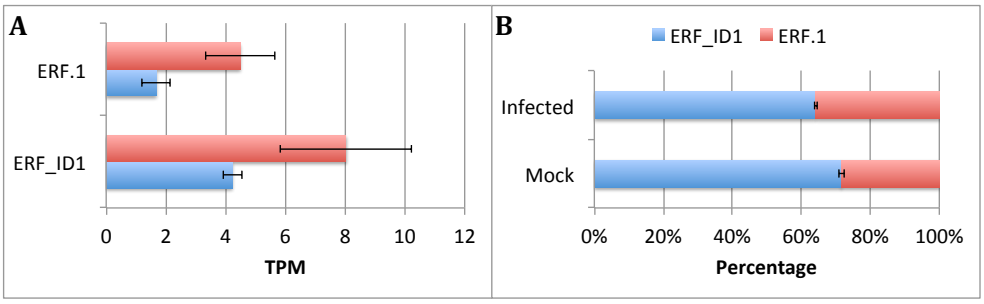


Figure 69. The proportion of *ERF.1* increases in response to *B. cinerea* 24 hpi. A. Transcript abundancies in TPM. Blue and red bars represent mock and infected samples. B. Ratio of isoforms in mock and *B. cinerea* infected Arabidopsis leaves. A and B. Error bars represent SEM of three biological samples.

We propose that in response to infection *ERF* undergoes DE, increasing the amount of both transcripts. In addition DAS occurs, resulting in the isoform containing the

tandem leucine rich repeats to increase in proportion relative to the isoform that lacks them.

DAS of *PEPR1* also results in a proportional decrease of the AS isoform that lacks the tandem leucine rich repeats (*PEPR1_ID2*) compared to the isoform that contains them (*PEPR1.1*). This potentially indicates that this is a general mechanism for regulation of the LRR-RLKs gene family.

Taken together these results indicate that there is a functional role for DAS in two unrelated important defence-related gene families, the RLKs and the JAZs in the *B. cinerea* defence response. Both of these share common patterns of how DAS can potentially affect gene and protein levels. Indicating that DAS is a biological mechanism for modulating and in part regulating the plants defence response.

To determine whether the *B. cinerea*-mediated DAS is specific to this biotic stress or is a more general stress response, our *B. cinerea*-mediated DAS gene dataset is compared to genes determined to undergo DAS in response to *Pst* and NaCl (Ding *et al*., 2014; Howard *et al*., 2013).

4.3.9.3 The same gene can undergo DAS in response to multiple stresses

Comparing *B. cinerea*-mediated DAS genes with genes DASed in response to two other stresses, *Pst* infection and NaCl, representing a hemi-biotrophic pathogen and abiotic stress respectively, shows that the same genes can undergo DAS in response to a variety of stresses, whilst others appear to be only DAS in response to *B. cinerea*. Howard *et al* (2013) identified 104 candidate *Pst*-mediated DAS Arabidopsis genes 12 hpi, with 29 of these genes having the isoform mixture changed by more than 10% in response to the *Pst* treatment. Ding *et al* (2014) identified 1636 DAS genes in Arabidopsis seedlings treated with 50, 150 or 300 mM NaCl. Comparing these to the 7578 *B. cinerea*-mediated DAS gene that we identified, there are 26 genes that are DAS by all three stresses, 49 DAS by both *Pst* and *B. cinerea*, 9 by both *Pst* and salt and 898 by both salt and *B. cinerea*. The majority of the *B. cinerea*-mediated DAS genes are unique to this stress and not found to be DAS by the other stresses (6605 out of 7578) Figure 70.

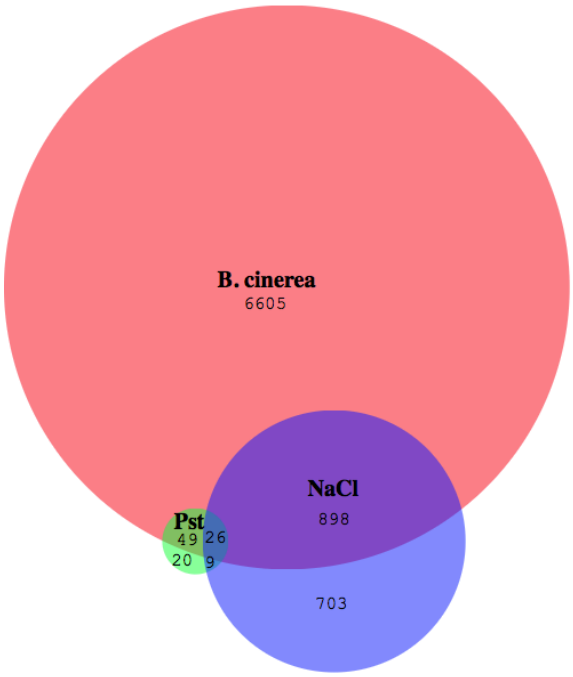


Figure 70. Comparison of DAS genes detected in response to three stresses. DAS in response to three stresses are compared: *B. cinerea* infection (red circle), *Pst* infection (green circle) and NaCl treatment (blue circle). The largest number of genes DAS in response to a stress are detected in response to *B. cinerea*, followed by NaCl with only relatively few genes being detected to undergo *Pst*-mediated DAS.

Our data utilised the reference transcript dataset, AtRTD, constructed by Zhang *et al* (2015), whilst the other two datasets used TAIR10 as the reference transcript. This could potentially result in us detecting additional DASed genes. However, of the 6605 genes that are detected to be DASed in response to *B. cinerea* and not the other stresses, only 11% of these originated from isoforms that are present in the AtRTD transcript set and not TAIR10. This indicates that this does not completely account for the larger number of DASed genes detected. The increased number of DASed genes detected in our dataset could be due to either our analysis not being stringent enough, or that the analysis carried out by Howard *et al* (2013) and Ding *et al* (2014) were not detecting all genes that undergo DAS in response to their respective stresses, i.e. they have many false positives. Both of these datasets have fewer replicates than our data (one per NaCl and two per *Pst* treatment). Fewer biological replicates could result in lower power for determining DASed genes. In addition, they were only analysed at one level, *Pst* was analysed at the isoform level and NaCl was analysed at the event level. As we determined earlier more than one

level of analysis appears to be required to effectively identify DASed genes. Thus the differences in the overall numbers could be due to the different analytical techniques. Supporting evidence for this is that the DAS dataset determined using event level analysis (the NaCl dataset) is larger than the dataset that utilised isoform level analysis (the *Pst* dataset). We identified more DASed genes at the event level compared to the isoform level, indicating that the difference in the number of DASed genes detected could be due to the RNA-Seq analysis method used. In particular, the *Pst* data, which used an isoform level analysis, only detects a very few isoforms that undergo DAS (104). We identified relatively few DASed genes utilising isoform level analysis. Only 24.4% of the 7578 genes undergoing DAS in response to *B. cinerea* were detected utilising isoform level analysis, indicating that the *Pst* dataset is not an exhaustive list of *Pst*-mediated DASed genes. Because there are likely to be more genes DASed in response to *Pst* and NaCl than in the datasets provided identifying stress specific DAS events from this information is unlikely to be accurate. However, investigating genes that are DASed in response to more than one stress may help to understand stress-mediated DAS.

Of particular interest are the 26 genes that undergo DAS in response to all three stresses; these could represent a mechanism for splicing being involved in a general stress response. Interestingly if only one level of analysis for *B. cinerea*-mediated DAS had been utilised not all 26 of these genes would have been detected. Fifty per cent of these 26 genes were uniquely identified by one level of analysis. The isoform, event and SJ level uniquely detected 11.5%, 23.1% and 15.4% respectively, with the remaining 50% being detected by more than one level of analysis. This adds weight to the idea that multiple levels of analysis should be utilised when investigating RNA-Seq data to determine DAS genes.

These 26 genes cover a range of biological functions including those involved in the splicing process, such as those that encode SR and SR like proteins (*ATSR30* and *SR45a*), genes whose expression increases in response to stress, for example CBL-Interacting Protein Kinase 3 *CIPK3*, and genes involved in regulating plant defence responses, for example CAM-Binding Protein 60-LIKE G (*CBP60G*) (Table 21). This

indicates that these 26 genes could be DASed as part of a generalised stress response, with genes involved in the regulation of splicing, plus those more directly involved in the stress response undergoing stress-mediated DAS.

Table 21. Genes DASed in response to NaCL, *Pst* and *B. cinerea* treatment. Column one gives the unique gene identifier number with gene names being found in column two. Column three gives a brief description of the gene (descriptions from the TAIR10 database, www.arabidopsis.org).

Identifier	Gene name	TAIR description
AT1G01550	BPS1, BYPASS 1	Encodes a protein with no functionally characterized domains that thought to prevent the synthesis of a novel substance that moves from the root to the shoot, where it modifies shoot growth by interfering with auxin signalling. Synthesis and delivery of this substance requires neither phloem nor endodermis.
AT1G03250		unknown protein
AT1G07350	SR45A	Encodes a serine/arginine rich-like protein, SR45a. Involved in the regulation of stress-responsive alternative splicing.
AT1G09140	ATSR30	Encodes a serine-arginine rich RNA binding protein involved in regulation of splicing (including splicing of itself). Exists as 3 alternative spliced forms that are differentially expressed.
AT1G10390	DRA2, DRACULA2	Nucleoporin autopeptidase
AT1G10590		Nucleic acid-binding, OB-fold-like protein
AT1G19660	ATBBD2	Wound-responsive family protein
AT1G69252		other RNA
AT1G75380	ATBBD1	Encodes a nucleases AtBBD1 involved in ABA-mediated callose deposition.
AT1G76990	ACR3, ACT DOMAIN REPEAT 3	ACT domain repeat 3
AT2G17500	PILS5, PIN-LIKES 5	Auxin efflux carrier family protein
AT2G26980	CBL-INTERACTING PROTEIN KINASE 3, CIPK3, SNF1-RELATED PROTEIN KINASE 3.17, SNRK3.17	encodes a serine-threonine protein kinase whose expression increases in response to abscisic acid, cold, drought, high salt, and wounding conditions. The gene is expressed in developing seeds and seedlings. Lines carrying a T-DNA insertions have reduced germination efficiency and expression of cold, high-salt, and abscisic acid marker genes are altered, but not drought-response markers.
AT2G31350	GLX2-5, GLYOXALASE 2-5	Encodes a mitochondrial glyoxalase 2 that can accommodate a number of different metal centers and with the predominant metal center being Fe(III)Zn(II).
AT2G38170	ATCAX1, CATION EXCHANGER 1, CAX1, RARE COLD INDUCIBLE 4, RCI4	Encodes a high affinity vacuolar calcium antiporter. The residue His 338 is critical to Ca2+ transport activity. Disruption of CAX1 reduces manganese and

		zinc of shoot tissue and results in a decrease in the activity of vacuolar V-type proton ATPase.
AT2G45170	ATATG8E, ATG8E, AUTOPHAGY 8E	Involved in autophagy. Under nutrient starvation the protein localizes to autophagosomes.
AT2G31060	EMB2785, EMBRYO DEFECTIVE 2785	Elongation factor family protein; Functions in: GTP binding, translation elongation factor activity, GTPase activity.
AT3G16400	ATMLP-470, ATNSP1, MYROSINASE-BINDING PROTEIN-LIKE PROTEIN-470, NITRILE SPECIFIER PROTEIN 1, NSP1	Encodes a nitrile-specifier protein NSP1 responsible for constitutive and herbivore-induced simple nitrile formation in rosette leaves. NSP1 is one out of five (At3g16400/NSP1, At2g33070/NSP2, At3g16390/NSP3, At3g16410/NSP4 and At5g48180/NSP5) A. thaliana epithiospecifier protein (ESP) homologues that promote simple nitrile, but not epithionitrile or thiocyanate formation.
AT3G24170	ATGR1, GLUTATHIONE-DISULFIDE REDUCTASE, GR1	Encodes a cytosolic glutathione reductase.
AT4G21850	ATMSRB9, METHIONINE SULFOXIDE REDUCTASE B9	methionine sulfoxide reductase B9
AT4G32440		Plant Tudor-like RNA-binding protein
AT4G35770	ARABIDOPSIS THALIANA SENESCENCE 1, ATSEN1, DARK INDUCIBLE 1, DIN1, SEN1, SENESCENCE 1, SENESCENCE ASSOCIATED GENE 1	Senescence-associated gene that is strongly induced by phosphate starvation. Transcripts are differentially regulated at the level of mRNA stability at different times of day. mRNAs are targets of the mRNA degradation pathway mediated by the downstream (DST) instability determinant.
AT4G39100	SHL1, SHORT LIFE	Putative transcription factor containing a PHD finger and BAH motif, required for normal development
AT4G39260	ATGRP8, GLYCINE-RICH PROTEIN 8 CCR1, COLD, CIRCADIAN RHYTHM, AND RNA BINDING 1,	Encodes a glycine-rich protein with RNA binding domain at the N-terminus. Protein is structurally similar to proteins induced by stress in other plants. Gene expression is induced by cold. Transcript undergoes circadian oscillations that is depressed by overexpression of AtGRP7. A substrate of the type III effector HopU1 (mono-ADP-ribosyltransferase).
AT5G16150	GLT1, GLUCOSE TRANSPORTER 1	Encodes a putative plastidic glucose transporter.
AT5G26920	CAM-BINDING PROTEIN 60-LIKE G, CBP60G	Encodes a calmodulin-binding protein CBP60g. The calmodulin-binding domain is located near the N-terminus; calmodulin binding is dependent on Ca(2+). Inducible by both bacterial pathogen and MAMP (microbe-associated molecular pattern) treatments. Bacterial growth is enhanced in cbp60g mutants. cbp60g mutants also show defects in salicylic acid (SA) accumulation and SA signalling.
AT5G59830		unknown protein

Comparing the two biotic stresses *Pst* and *B. cinerea* could help to determine if there are common genes that undergo DAS as part of a defence related mechanism. There

are 75 genes that undergo DAS in response to both the hemibiotrophic pathogen *Pst* and the necrotrophic pathogen *B. cinerea*. These 75 genes are enriched for defence related GO terms, for example “response to biotic stimulus”, and “immune response” compared to the genome as a whole. They are also enriched for the GO terms associated with other stress responses such as “response to abiotic stimulus” (17 genes) and “response to salt stress” (9 genes) (Figure 71). The majority of the genes in each of these clusters were not detected to be DASed in the NaCl dataset (10 of the 17 genes in the response to abiotic stimulus cluster and six of the nine genes in the response to salt stress cluster), potentially indicating that these genes detected to undergo DAS in response to both *B. cinerea* and *Pst* may also be involved in a generalised stress response.

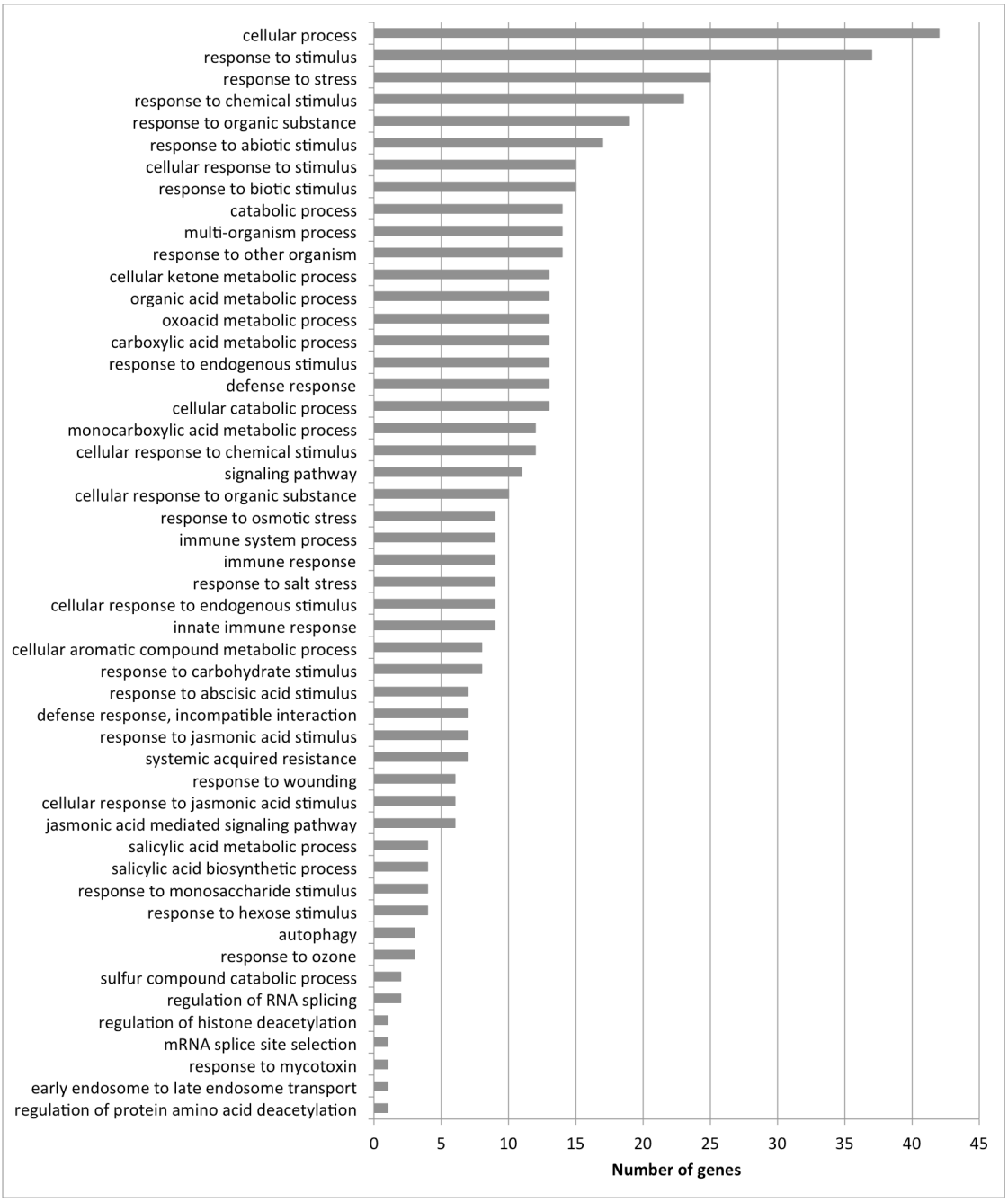


Figure 71. Genes DASed in response to both *Pst* and *B. cinerea* infection are enriched for defence related GO terms. Grey bars represent the number of genes belonging to a certain GO category. GO enrichment terms determined using BINGO 3.0.3 and are organised by number of genes per cluster (largest at the top) (Maere *et al*., 2005).

Of the 75 genes that undergo DAS in response to *B. cinerea* 24 hpi and *Pst* 12 hpi, 16 of these were also deemed to be DEG in response to *B. cinerea* in section 4.1.1.3 page 134. 24 of these also being shown to be DE in response to *Pst* treatment by Howard *et al* 2013. Ten genes were DE in response to both *Pst* and *B. cinerea*, with

the remaining 6 and 14 genes only being DE in response to *B. cinerea* and *Pst* infection respectively. This indicates that both DE and DAS may contribute to the plant defence responses.

The genes that are not DE in response to *B. cinerea* 24hpi but do undergo DAS include some genes known to be involved in plant defence response. These include JASMONATE RESISTANT 1 (*JAR1*); loss-of-function mutants of *jar1* are more susceptible to *B. cinerea*, and NINJA, which is involved in regulating JA responses (Ferrari, Plotnikova, De Lorenzo, & Ausubel, 2003; Pauwels *et al*., 2010). In addition, they also contain genes involved in regulating splicing such as the SR protein *AtSR30* and the hnRNP protein, Glycine-Rich Protein 8 (*AtGRP8*). This indicates that there is a putative subset of functionally significant genes that could be regulated by DAS in response to stress. Interestingly for one of these genes (*AtSR30*) the isoform that is detected to be differentially expressed differs between the *Pst* and *B. cinerea* experiments. However, as previously stated we expect there to be more *Pst*-mediated DASed genes than detected by Howard *et al* (2013). Thus to help determine if plants response to different stresses utilise different isoforms comparable experiments with similar analysis would be required.

Taken together our experiments indicate that DAS can occur as a response to stress, affecting gene expression levels and functional domains. Now we begin to investigate potential mechanisms that could regulated stress responsive DAS.

4.3.10 Beginning to elucidate regulatory mechanisms of defence related DAS

Regulation of splicing is a complex process that has been shown to involve cis-acting splicing regulatory elements (SREs) that bind to trans-acting factors such as serine arginine rich (SR) proteins and the heterogeneous nuclear ribonucleoprotein (hnRNP)-like proteins. To being to elucidate the mechanisms behind regulation of *B. cinerea*-mediated DAS, splicing factor genes will be investigated to help determine which trans-acting factors are affected, and motif analysis will be carried out to identify potential defence associated cis-acting SREs.

4.3.10.1 B. cinerea-mediated DAS and DE genes are enriched for splicing associated GO terms

Splicing factor genes such as those that encode SR proteins and hnRNPs are often regulated by AS and gene expression processes. Functional analysis utilising GO enrichment of our *B. cinerea*-mediated DAS and DEG datasets shows that RNA processing terms, including those linked to splicing are enriched. This suggests that one of the underlying biological process affected by both *B. cinerea* mediated DAS and DGE may be associated with splicing factor genes. To investigate this further a list of 224 known splicing factors associated genes was compiled and investigated to see which genes were DE and/or DAS in response to *B. cinerea* 24 hpi.

4.3.10.2 Splicing factors are differentially alternatively spliced in response to B. cinerea infection

Splicing factors undergo DAS in response to *B. cinerea* 24 hpi but only a few of the splicing factor genes are DE. Approximately half of the splicing factor associated genes (118/224) undergo *B. cinerea* mediated DAS, whereas only eight splicing factor genes are DE; four of these also undergo DAS in response to *B. cinerea* 24hpi. This indicates that splicing factor genes are predominantly regulated by DAS.

One example of this is the functionally redundant components *mac3a* and *mac3b* of the MOS4 associated complex (MAC). Monaghan *et al* (2009) have shown that MAC3A and MAC3B are functionally redundant in the regulation of plant innate immunity, and we have previously shown in Arabidopsis plants that the double loss-of-function mutant *mac3a mac3b* is more susceptible to *B. cinerea* infection (Foster, 2012). The *mac3a mac3b* double loss-of-function mutants alters the splicing patters of *SNC1* and *RPS4* (F. Xu *et al*., 2012). Thus, these genes have been linked to both regulating plant defence and altering splicing patterns of defence-related genes. Because *MAC3A* and *MAC3B* do not undergo DE at the gene level, but are DAS in response *to B. cinerea* this could potentially indicate that DAS of some splicing factor associated genes could affect the splicing pattern of downstream defence-related genes. Investigations into the *MAC* genes and how they are affected by and affect DAS of other genes may help to elucidate how splicing factors influence the

interwoven network of splicing genes that interact with differentially expressed gene networks. Further work on this subject is carried out in “Chapter 5: The spliceosome and the defence response; investigating the MOS4 associated complex.”

Evidence that some splicing factor genes are predominantly regulated by DAS comes from investigating the SR gene family. The percentage of SR genes that are DAS in response to *B. cinerea* (77.8%) is significantly higher compared to all splicing factor associated genes (exact binomial p value = 0.03), and none of these SR protein encoding genes are DE in response to *B. cinerea* (Figure 72). The exact binomial test was used because we have discrete data that falls into two categories (SR genes and all splicing factor associated genes) and we want to test whether the observed proportion of individuals in the all splicing factor associated gene category deviates significantly from the proportion observed in the SR genes.

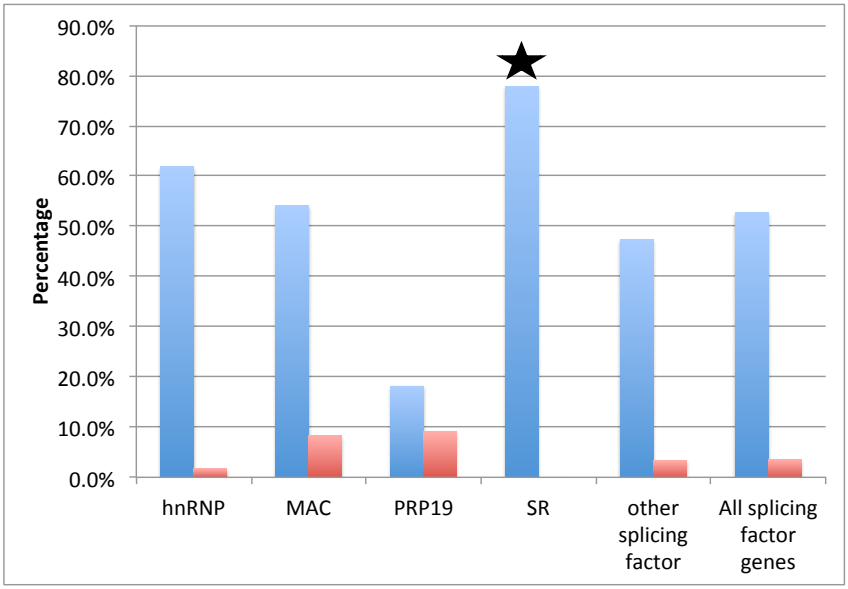


Figure 72. Approximately half of the 224 splicing factor associated genes undergo *B. cinerea*-mediated DAS. The SR subgroup has a statistically significant (indicated by higher percentage of genes that are DAS in response to *B. cinerea* compared to splicing factor associated genes as a whole (p value = 0.03 exact binomial test). Blue and red bars represent the percentage of gene that are DASed and DE respectively.

Mining the current literature for evidence of stress-related DAS of genes that encode the SR proteins show that they often undergo DAS in response to stress (Table 22).

Table 22. SR proteins are DASed in response to a variety of stresses in Arabidopsis. Column one and two detail the gene identifier and gene name respectively, with column three indicating under which stress the gene has been shown to undergo DAS. Green rows indicate that the gene undergoes *B. cinerea*-mediated DAS in our dataset. Evidence from DAS obtained from Lazar and Goodman 2000, Palusa *et al* 2007, Ding *et al* 2014, Howard *et al* 2013, and Filichkin *et al* 2010.

Gene Identifier	Gene Name	DAS in response to other stresses
AT1G02840	SR34/SR1	temperature, hormones, heat and cold
AT1G09140	SR30	salt, Pst, high light and heat stress
AT1G55310	SCL33	salt, ABA, IAA, BA, NaCl, cold, heat, glucose
AT2G24590	RSZ22a	-
AT2G37340	RSZ33	salt, heat, glucose
AT2G46610	RS31A	heat
AT3G13570	SCL30a	salt, heat
AT3G49430	SR34a	salt
AT3G53500	RSZ32	salt, heat
AT3G61860	RS31	salt, cold
AT4G02430	AtSR34b	Heat, cold, hormones and NaCl
AT4G25500	RS40/RSP35	salt, cold and heat
AT4G31580	RSZ22/SRZ-22	-
AT5G52040	RS41	salt, heat
AT1G23860	RSZ21/SRZ-21	-
AT3G55460	SCL30	salt
AT5G18810	SCL28	-
AT5G64200	SC35	-

AS mentioned earlier one of these genes *SR30* is DAS in response to both *B. cinerea* and *Pst* infection but different isoforms were detected to be DAS. Howard *et al* (2013) detected the isoform *SR30.2* to be DASed with this isoform being decreased in infection, whereas we detected the isoform *SR30.1* to undergo *B. cinerea*-mediated DAS with an increase in response to infection. To investigate this further, total transcript abundances and the proportional change of isoforms was investigated. In response to *B. cinerea* infection, *SR30.2* does not change in abundance whilst *SR30.1* increases. This results in a proportional change of 13% where *SR30.1* increases whilst *SR30.2* decreases (only two isoforms, *SR30.1* and *SR30.2* isoforms were detected) (Figure 73).

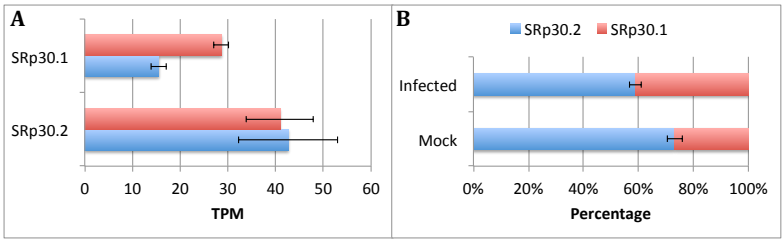


Figure 73. *SR30.1* increase in abundance response to *B. cinerea* infection. A. Transcript abundancies in TPM. Blue and red bars represent mock and infected samples. B. Ratio of isoforms in mock and *B. cinerea* infected Arabidopsis leaves. A and B. Error bars represent SEM of three biological samples.

The difference between these two isoforms is an A5SS event in the tenth exon that is likely to encode part of the C terminal domain. This AS event could thus result in proteins that have different C-terminal domain structures. The C terminal domains of SR proteins have been associated with recruiting other proteins, contacting the pre-mRNA branch point or containing signals for cellular localization (Caceres *et al*., 1998; Long & Caceres, 2009; H. Shen & Green, 2006). This indicates that in response to stress one or all of these functions may be affected.

Both isoforms of *SR30* contain two RNA recognition motifs (RRMs) that are unaffected by the DAS event. However, the RRM domains are affected by *B. cinerea*-mediated DAS in other SR proteins. For example *RS31* and *RS31a*. *RS31* and *RS31a* encode two members of the RS subfamily, which are plant specific SR proteins that contain two RRM domains. In response to *B. cinerea* infection, the isoforms that contain intact RRM domains decrease whilst those isoforms where an AS event results in disruption of the first RRM domain increase in proportion. Disruption of the RRM domain as a result of *B. cinerea* mediated DAS is also seen in SR protein with only one RRM domain, for example *SCL30a* where in response to infection the isoform that is predicted to encode the isoform with the intact RRM domain (*SCL30a.1*) decreases in proportion to *SCL30a_ID2* and *SCL30_ID4*, which are predicted to encode proteins with disrupted RRM domain (due to RI events in this region).

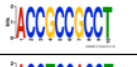
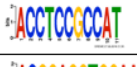
The RRM domains bind to splicing regulatory elements (SRE) in target genes, thus disruption of these domains is likely to affect the splicing of these target genes. To

help to elucidate if there are any candidate *B. cinerea* specific SREs, motif analysis is carried out.

4.3.10.3 Potential SRE motifs associated with B. cinerea-mediated RI events are detected

Examining the DNA architecture within RIs utilising Cooke (2013) dataset in Chapter 3 section 3.3.5 page 3.3.5 we identified some potential motifs as a candidate infection regulating splicing motifs. To further investigate potential SRE motifs in our dataset that contains biological replicates, we want to construct at a restricted dataset to minimise false positives, resulting in a higher-powered analysis. Again, our analysis is restricted to investigating intronic regions and determining if there are any motifs associated with *B. cinerea* mediated RI events. There was ambiguity of some genes being DE rather than undergoing DAS at the isoform and SJ levels of analysis. To minimise this, the intersect from all three analyses is utilised to provide a conservative dataset. This conservative dataset consists of 314 RI events, 205 that are up-regulated and 109 that are down-regulated in response to *B. cinerea* infection. To elucidate SRE motifs DREME on CyVerse was utilised to determine which motif were enriched compared to the TAIR10 RI set determined by Mao *et al* 2014 (Bailey, 2011; Goff *et al*., 2011; Mao, Raj Kumar, Guo, Zhang, & Liang, 2014), with a motif range of 3 to 12 nt. The up-regulated and down-regulated RI sequences were analysed separately with 22 and 24 significantly enriched motifs being identified respectively (Table 23).

Table 23. Motifs enriched in *B. cinerea*-mediated DAS RI. Columns 1 and 4 show the sequence logos of the motifs (columns 2 and 5) enriched in RI up-regulated and down-regulated in response to *B. cinerea* 24 hpi respectively (columns 3 and 6 give the unerased E values). The unerased E value is the E value of the motif that is calculated without erasing sites of previously matched motifs. The E value is the P value (calculated by Fisher's exact test) times the number of candidate motifs, Motifs obtained using the DREME software via CyVerse (Bailey, 2011; Goff *et al*., 2011).

Up regulated RI			Down regulated RI		
Sequence logo	Motif	Unerased E value	Sequence logo	Motif	Unerased E value
	GCNGM	2.3e-4		AWGGTGGTGG	2.7e-2
	CCDKC	3.4e-1		AGGCGGTGGA	4.8e-4
	GAGNC	6.0e-2		AGCTGGTGGRC	8.3e-4
	CSTGA	4.0e-4		AGGAGGATMT	8.3e-4
	GGTR	1.5e-4		AGGAGGYGGAGC	8.3e-4
	GGAAGY	7.3e-4		AWTGGTGGGG	8.3e-4
	CTCGWG	4.2e-5		ATGTCCCYCG	8.3e-4
	KCTCCK	3.4e-6		GCCTCCTCCA	8.3e-4
	AAGAGGKT	4.1e-3		CCTYCGCCAG	8.3e-4
	CCCDCG	1.3e-5		AAGGTGC	2.1e-3
	AATYATCAAG	5.0e-2		ATCCTCCAC	6.2e-3
	GTCCAY	3.2e-4		CCGTATCCM	6.2e-3
	ACTGG	7.7e-3		ACCTCCGKAAC	8.2e-4
	GAGGATS	6.4e-3		ACCGCCGCCT	1.2e-2
	CACAGTWA	1.2e-2		ACCTCCACCT	1.2e-2
	AAATHAATGG	9.6e-3		ACCTCCGCCAT	1.2e-2
	CGYGC	1.4e-1		ACGGAGGTGGAA	1.2e-2
	AGCTTTWCG	9.4e-3		AKGCGGA	3.2e-3
	AGGTCAAMG	3.6e-3		ACGGTGGTGGA	1.2e-2
	CCACTM	4.7e-5		AGGTGGTGGAC	7.9e-2
	GWCGM	1.8e-3		AGGTGGTGGCC	1.2e-2
	CCTTR	2.6e-1		CCATGCCCA	1.2e-2
				AGCGKTGA	2.1e-3
				GCATKAA	1.1e-2

The motifs enriched in up and down-regulated RI events differ in length and composition, with the up-regulated motifs being shorter on average compared to the down-regulated motifs. The difference in composition could indicate that different splicing factor genes are involved in binding intronic regions, resulting in *B. cinerea*-mediated up-regulated RI events, compared to those that are down-regulated.

Mining literature from a variety of species indicates that there are four RBP that bind to similar motifs to those found in *B. cinerea*-mediated up-regulated RI in humans (*SRp40*, *SRp20*, *SF2/ASF* and *RBM5*), and one that has previously been identified in plants (*SCL33/SCL30a*). This could potentially indicate that for those RIs that are up-regulated in response to *B. cinerea*, homologues of human genes may bind to these candidate SREs. *SRp40* that binds to the motif YYWCWSG, which resembles the CCACTM motif enriched in *B. cinerea*-mediated RI, and *SRp20* binds to CTCKTCY, which resembles our motif CTCGWG. There are no clear orthologues in Arabidopsis of *SRp40* and *SRp20*. However this could potentially indicate that another plant SR protein could bind to these regions.

In humans SF2/ASF binds to CRSMSGW, which resembles our motif CACAGTWA. Arabidopsis contains two SF2/ASF like proteins, *SR30* that is a splicing modulator and *SR34* that acts as a general splicing factor (Lopato *et al*., 1999). Binding experiments could inform if either of these two SR proteins bind to this candidate SRE. The RBP RBM5 in humans binds to TCATC, which occurs in our motif AATYATCAAG. This could be particularly interesting to investigate further because a plant homologue of RBM5 is SUA, which has been shown to be required for splicing of *CERK1*, (an important chitin receptor). This represents a potential link between a RBP associated with *B. cinerea*-mediated RI events and a gene with a known role in plant pathogen defence.

The plant SR protein *SCL33* and its close homologue *SCL30a* have been shown to bind to a GAAG motif. This motif is also detected to be enriched in genes with RI events that are up-regulated in response to *B. cinerea* 24hpi; it forms part of the motif GGAAGY. This purine rich motif, GAAG, has been implicated to act as a

regulators of AS that contributes to RIs in SR genes (Thomas *et al*., 2012). This could potentially indicate that this motif is involved in infection specific regulation of SR proteins involved in defence related DAS.

The enriched motifs in *B. cinerea*-mediated down-regulated RI have multiple motifs containing three main sequences GGTGGT, GGAGGA and ACCTCC. The GGAGGA motif has been linked to SRSF1 binding and the GGTGGT motif binds hnRNPM in humans. This indicates that both SR proteins and hnRNP proteins may be involved in DAS in response to *B. cinerea*.

The motif ACCTCC, which we saw in multiple enriched *B. cinerea*-mediated down-regulated RIs has been investigated by Culler *et al* 2010 in humans. This motif bears similarities to the PTB motif, but when Culler *et al* 2010 depleted PTB only a very small change was observed. However, when CUG-BP1 was depleted (which binds CTG and GT rich motifs) a greater than 2-fold decrease in exon exclusion was detected. These authors state that CUG_BP1 is unlikely to bind this motif but may be recruited through interaction with as yet unknown proteins. This could indicate that similar mechanisms are involved in the regulation of *B. cinerea*-mediated DAS.

4.4 Discussion

4.4.1 Multiple pipelines provide a more comprehensive DAS dataset

We found that identifying DAS using three levels of analysis resulted in unique genes being detected at each level. This indicates that multiple levels of analysis of RNA-Seq data will provide a more comprehensive DAS dataset. However, there are some major drawbacks to this approach, firstly it has the potential to increase the false positive rate, and secondly it can dramatically increase the computational and analytical time required.

Literature shows that different methods for detecting DAS events results in very little overlap between them. Liu *et al* 2014 compared six well known methods, DiffSplice, CuffDiff, DEXSeq, MATS, rDiff-Param and Splicing Compass, and determined that at least 50% of DASed genes detected by each method were uniquely detected by that method (Ruolin Liu *et al*., 2014). Our different methods also showed a relatively small overlap. The event level analysis detected the largest number of total events and also had the largest percentage of uniquely identified DAS. Although the SJ and isoform level approach detected a relatively large number of unique genes, due to them detecting considerably less DAS genes this actually represents only a small proportion of the dataset, 17.6% and 7.19% respectively. The uniquely detected DASed genes were not over-represented or under-represented for any GO terms compared to the DAS dataset as a whole and did not comprise of any specific gene family, indicating that the each of the techniques will provide information on potential functional roles and gene families affected.

4.4.2 The number of RNA-Seq analytical methods used depends on the biological question asked

The decision on what analysis to use very much depends on the biological question being considered. For example, to compare stresses and identify unique and hub DASed genes, an exhaustive list may be required that detect as many genes as possible. In this case, all three methods of analysis would be the appropriate approach. However, if a general overview of differences of DAS between treatments

or lines is required an event level only approach may be more suitable. This is because it is likely to detect the majority of DASed genes and will take considerably less computational and analytical time. This chapter utilised two approaches: firstly we created a relaxed dataset in order to investigate how widespread *B. cinerea*-mediated DAS was, and to compare *B. cinerea*-mediated DAS with other stress responses because for this analysis we wanted to maximise our chance of detecting DASed genes. Secondly for detecting candidate SRE motifs we used a conservative dataset consisting of the intersect of the three methods. A conservative dataset was utilised here because we were interested in detecting patterns of nucleotides rather than detecting individual genes and wanted to minimise the number of false positives to make the motif searching more powerful.

One important thing to note is that because different methods have very little overlap in detecting DAS, great care needs to be taken when comparing different datasets. This is because the differences in DASed genes may be due to the method used to determine them rather than due to biological significance.

4.4.3 Arabidopsis leaves infected with *B. cinerea* undergo genome-wide differential alternative splicing

B. cinerea-mediated DAS occur in 34% of Arabidopsis intron-containing genes at 24 hpi, with these genes being enriched for GO terms associated directly with defence responses and splicing terms, as well as some terms that although not directly defence related have been shown to be affected in response to stress, such as those associated with metabolism, regulation of biological processes, and transport. This indicates that infection-responsive DAS plays a functional role in the defence response.

Although the patterns of how exons and introns combine changes in *B. cinerea* infected plants, the proportions of event types seen to be AS is roughly the same as those seen to be DASed, indicating that no one event type is preferred under *B. cinerea* stress conditions. To begin to elucidate how *B. cinerea*-mediated DAS may play a functional role in the defence response and its potential changes to the

proteome two distinct defence-related gene families were looked at in more details, the RLK and the JAZ families.

4.4.4 *B. cinerea*-mediated DAS has the potential to alter proteins concentration, localisation and functional domains

It has long been established in eukaryotes that AS can cause one gene to encode protein isoforms with different domains, activity or localisation depending on which transcript is produced (Stamm *et al*., 2005). Relatively recently this has also been shown in plants. In addition it has been demonstrated to be involved in the fine tuning of protein expression levels (Kalyna *et al*., 2012; Syed *et al*.,2012). Utilising the nucleotide sequences of *B. cinerea*-mediated DASed genes, we predict that infection responsive DAS could cause each of these phenomena. From our analysis of the two defence-related genes families, two main potential methods for *B. cinerea*-mediated DAS affecting protein levels were evident; PTC coupled to NMD and varying 5'UTR length.

4.4.5 B. cinerea-mediated DAS potentially utilises premature termination codons linked to nonsense-mediated decay to alter protein levels

Drechsel *et al* (2013) determined that approximately 18% of Arabidopsis genes that are AS produce mRNA transcripts that are targets for the NMD machinery. This mechanism has the potential to regulate the amount of protein produced by a gene via modulating transcript levels. In both the JAZ and RLK families, examples of *B. cinerea*-mediated DAS genes containing isoforms where PTCs are likely to trigger NMD are detected. This indicates that one of the potential effects of *B. cinerea*-mediated DAS is to modulate transcript levels, thus affecting the protein expression levels. This is in line with current literature where Gloggnitzer *et al* (2014) have shown that AS coupled to NMD regulates Arabidopsis response to the bacterial pathogen *Pst*. Thi,s combined with our work indicates that this is a common mechanism found in plants in response to stress. Experimental data focusing on a group of rice dehydration-responsive element-binding protein 2s (DREB2s), showed AS coupled to NMD can induce a rapid response to stress resulting in increased

protein levels that does not require transcriptional activation (Matsukura *et al*., 2010; Staiger & Brown, 2013). AS of *DREB2B* produces two different isoforms; *DREB2B1,* which due to a PTC is non-functional, and the functional isoform *DREB2B2.* Stress-inducible AS causes an increase of the functional isoform and thus, increases protein levels independently of transcriptional activation (Matsukura *et al*., 2010; Staiger & Brown, 2013). Putative examples of this rapid response mechanism are found in *B. cinerea*-mediated DAS in both JAZ and the LRR-RLKs gene families. Our RNA-Seq experiment detects four isoforms of *JAZ9*; two functional isoforms *JAZ9.1* and *JAZ9.2* and two isoforms that due to PTC are likely to be targeted for NMD, *JAZ9_ID1* and *JAZ9_ID2*. Under *B. cinerea* infection conditions the levels of the non-functional transcripts decrease indicating that protein levels will increase due to a transcriptionally independent mechanism. This also occurs in the RLKs family, for example in the gene that encodes one of the major modulators of plant PTI *BAK1*. The non-functional isoform *BAK_ID3* decreases by 18% in *B. cinerea* infected leaves whilst the functional *BAK1.1* increases, once again indicating that infection with *B. cinerea* could lead to an increase in protein via transcriptional independent means. Interestingly Nolan *et al* (2011) proposed that in *M. truncatula MtSERK3* (*BAK1* is also known as *SERK3*) could undergo defence related AS, although this has not been experimentally verified, potentially indicating that DAS of *SERK* genes may play a role in plant defence across a range of species.

From these experiments we propose that one mechanism of regulating the plant defence response is through DAS coupled to NMD, which enables a rapid response to infection by circumventing the time required for transcriptional activation and pre-mRNA accumulation although further experimental work would be required to confirm this.

4.4.6 Differential alternative splicing of 5'UTRs is predicted to be a mechanism for altering protein levels in response to infection

There is a direct correlation between the length of introns present in the 5'UTR and the efficiency of gene expression levels, with the longer the intron the more the gene is expressed (B. Y. W. Chung *et al*., 2006). However, this paper did not

investigate if uORF are present nd weather this coupled to nonsense mediated decay could be the mechanism behind the altered gene expression. Both of these potential mechanisms could affect the amount of protein produced, and thus DAS that alters the length of introns in 5'UTR regions could potentially alter the levels of proteins produced. DAS associated with changing the length of 5'UTRs occurs in both the JAZ and the RLKs families.

The example from investigating the JAZ family comes from *JAZ3*. *JAZ3* is not DE in response to *B. cinerea* but has two isoforms that show significant changes in their proportions in infected samples compared to mock samples. In infected samples, the proportion of *JAZ3.3* is increased, accompanied by a decrease in *JAZ3.2*. Both of these isoforms have the same coding region with the differences between the isoforms occurring in the 5'UTR. *JAZ3.3* contains an intron, which is retained in *JAZ3.2* resulting in *JAZ3.2* having a longer 5'UTR, potentially increasing the amount of protein produced. However multiple AUG-initiated uORF are also detected in this isoform potentially making it a target for NMD mechanism.

Another hypothesised mechanism whereby the RI event in the 5'UTR of *JAZ3.2* could increase the amount of the JAZ3 protein in mock samples comes from work carried out by Remy *et al* (2014) on zinc (Zn) tolerance in Arabidopsis. Remy *et al* 2014 showed that ZIF2 (Zinc-Induced Facilitator 2) is important for Zn tolerance. This Zn-induced gene can produce two isoforms with the same coding area but ZIF2.1 has an intron which is retained in the AS isoform ZIF2.2(Remy *et al*., 2014). The RI event in ZIF2.2 5'UTR enhanced translation in a zinc responsive manner, significantly increasing the amount of ZIF2 protein produced in excess Zn conditions. The enhanced translation is however likely due to a predicted stable stem loop immediately upstream of the start codon that is located in the RI and thus not present in the *ZIF2.1* 5'UTR rather than due solely to an increase in the length of the 5'UTR (Remy *et al*., 2014).

In the example of ZIF2, the isoform with the RI increased in response to stress thus increasing the protein levels. Investigating the RLKs identified a gene where this could be the case in our data. The RLK encoding gene *PBL27* also produces two

isoforms, *PBL27.1* and *PBL27.2*, where the coding regions are the same, but there is a difference in the length of the 5'UTR due to a RI in the *PBL27.2* isoform. In infected samples the levels of *PBL27.2* increases by 23.4% increase. Thus we predict that in response to *B. cinerea* infection the isoform with the RI increases resulting in enhanced translation and thus protein levels. This would indicate that protein regulation could occur via DAS being linked to changes in secondary structures of RNA that affect the efficiency of translation. However the RI event in PBL27.2 results in a GUG-initiated uORF thus potentially targeting it for NMD.

We hypothesise that *JAZ3* and *PBL27* have two isoforms associated with stress-induced DAS, which can alter protein levels. However further work would be required to confirm if protein levels are altered in infected samples and to determine how the DAS events cause this. Two potential mechanisms are DAS coupled to NMD, and alterations of the RNA secondary structure. Changing the RNA secondary structures (due to RI in the 5'UTR) has the potential to enhance protein translation in the RI isoform, and thus, increase protein levels. Stress-mediated DAS can alter the isoforms percentages and thus alter the protein levels; in the case of JAZ3 protein levels would decrease due to *B. cinerea*-induced DAS reducing the amount of the RI isoform *JAZ3.2*. Whereas in PBL27 the protein levels would increase due to an increase in the RI isoform (*PBL27.2*). Further work would be required to test this hypothesis, including but not limited to validating the *B. cinerea*-induced DAS, and investigations into both protein levels and RNA secondary structures. If this hypothesis was determined to be valid a genome-wide investigation of RI events in 5'UTRs due to a variety of stress-induced DAS could determine if this is a generalised response to stress.

In addition to the above-proposed roles for *B. cinerea*-mediated DAS altering protein levels, it is also predicted to affect their functional domains.

From investigations into the JAZ and RLKs families two main effects of DAS altering functional domains are predicted; truncated proteins acting as dominant-negative inhibitors of full length proteins, and loss of functional domains potentially resulting in altered binding specificities.

4.4.7 *B. cinerea*-mediated DAS is predicted to result in altered levels of truncated proteins that are potential components of regulatory circuits

Seo *et al* (2011) identified an AS isoform of the transcription factor *INDERMINATE DOMAIN 14* (*IDD14*), *IDD14β*, that lacks a functional DNA-binding domain and is predominately produced in cold conditions. This isoform can form heterodimers with the functional form *IDD14α* and these have reduced binding activity to the Qua-Quine Starch gene that regulates starch accumulation. Thus they proposed that proteins with truncated domains as a result of AS can create a regulatory loop that may modulate starch accumulation in response to cold (Seo *et al*., 2011). Chung *at el* (2010) proposed a regulatory loop for JA signalling that can be modulated by AS of *JAZ10*. They demonstrated that different isoform of *JAZ10* had different affinities of binding to COI1 due to production of proteins with truncated Jas domains, resulting from all or partial retention of the Jas intron (an intron within the area of DNA that encodes the jas domain). They demonstrated that both *JAZ10.3* and *JAZ10.4* have reduced capacity to form stable complexes with COI1; in *JAZ10.4* isoforms this ability is abolished whereas in *JAZ10.3* it is significantly reduced. *JAZ10.1* strongly binds COI1 in the presence of JA-Ile, but is less stable than *JAZ10.3* and *JAZ10.4*, thus *JAZ10.3* and *JAZ10.4* act as dominant repressors of JA signalling (H. S. Chung *et al*., 2010). Recently Moreno *et al* (2013) added to this body of work by demonstrating that *JAZ10.4* could act within this circuit to provide negative feedback. In response to *B. cinerea* infection, the levels of JA-Ile increase resulting in the formation of COI1-JAZ co-receptor complexes, which target JAZ proteins for SCF^{COI1} mediated ubiquitination, and thus release transcription factors from inhibition, leading to activation of JA-responsive genes (Chini *et al*., 2007; Pauwels *et al*., 2010; Thines *et al*., 2007; Yan *et al*., 2007). *JAZ10.3* and *JAZ10.4* could act as dominant repressors restraining this activation of JA-responsive genes, to potentially circumvent the fitness cost associated with inappropriate or hyper activation of these responses, thus creating a regulatory circuit. In our RNA-Seq dataset, the predominant proportional change in isoforms is associated with the dominant repressor isoforms; *JAZ10.4* decreases in proportion and *JAZ10_ID3* increases in proportion in *B. cinerea*

infected samples. We hypothesise that this *B. cinerea*-mediated DAS could modulate the JA signalling pathway by altering the component that is restraining the activation of JA-responsive genes. We propose that in mock samples *JAZ10.4* strongly represses the activation of JA-responsive genes to prevent inappropriate activation of these genes. In response to *B. cinerea* infection, this restraint needs to be lifted but not completely abolished because mechanisms are required to prevent hyper activation. Thus the *JAZ10.3* isoform is utilised because it still represses the system but to a lesser extent than *JAZ10.4*.

In addition to altering protein domains we detected some *B. cinerea*-mediated DAS that is predicted to completely remove certain domains of proteins.

4.4.8 *B. cinerea*-mediated DAS alters the tandem leucine rich repeats of leucine rich repeat receptor like kinases

We have discovered two examples of where *B. cinerea*-mediated DAS reduces the proportion of LRR-RLK isoforms that are predicted to lack the LRR tandem repeats. *EFR* has two isoforms, *EFR.1* that encodes the full protein and *EFR_ID1* that lacks the tandem LRR due to an A5SS. In infected samples, a lower proportion of *EFR_ID1* is detected. A similar scenario occurs with *PEPR1* whereby the isoform that is predicted to lack the tandem LRR (due to an additional intron in this region). *PEPR_ID2* is also found in reduced proportion in infected samples. Drawing on work carried out in R genes by Meyers *et al* (2002) who proposed that R proteins lacking leucine rich repeats domains may act as adapter proteins, we hypothesise that this may also be the case for LRR-RLKs. We propose that AS of LRR-RLK genes can cause some isoforms to lack the LRR domains. These isoforms may be acting as an adapter protein to help mediate signal transduction with its expression fine-tuned via DAS; this would however need to be experimentally validated.

4.4.9 Splicing factors are differentially spliced in response to *B. cinerea* infection.

Our data indicates that the majority of splicing factor associated gene families are regulated at least in part by DAS. Each of the splicing associated gene families contained some genes that underwent *B. cinerea*-mediated DAS but are not DEGs.

This is most predominant in the SR protein family where 77.8% of this family undergo DAS but none are DEGs.

It is well known that members of the SR protein family can be regulated by DAS under a variety of different stresses, with the majority of them being DAS under multiple stresses (see Table 22 for details) (Ding *et al*., 2014; Filichkin *et al*., 2010; Howard *et al*., 2013; Lazar & Goodman, 2000; Palusa *et al*., 2007). Noticeable exceptions are RSZ21, SCL28 and SC35, which to our knowledge have as yet not been detected to be DASed in response to any stress, and we did not detect them to undergo *B. cinerea*-mediated DAS. However, we detected two SR proteins that to the best of our knowledge, have not previously been detected to undergo DAS in response to stress in Arabidopsis; RSZ22 and its paralog RSZ22a. However the *Arachis diogi* gene RdRSZ1, which shares 78.5% sequence homology to RSZ22, is induced in response to the fungal pathogen *Phaeoisariopsis personata* and is implicated to play a role in plant death and HR-like cell death (Kumar & Kirti, 2012). Kumar & Kirti (2012) determined that there are two isoforms of RdRSZ1 that differ in the 3'UTR. They state that the differences in the 3'UTR played no direct role in the observed defence phenotype but may play a role in the stability of the transcripts (Kumar & Kirti, 2012). An altered 3'UTR of AtRSZ22 appears to be associated with the *B. cinerea*-mediated DAS A5SS event seen in our dataset. However, further work would be required to determine this, potentially indicating that *B. cinerea*-mediated DAS affects the stability of AtRSZ22 transcripts. The fact that there is a close relation to RSZ22 that is induced in response to a fungal pathogen which appears to pay a role in plant defence could indicate that the SR protein, RSZ22, and its paralog, RSZ22a, are specifically induced in response to biotic pathogens. Further evidence for this is that Palusa *et al* (2007) carried out a detailed screen of tissue type, hormones and abiotic stresses, none of which resulted in AS of AtRSZ22. In particular, this SR protein may specifically be associated with fungal pathogens because Howard *et al* (2013) did not detect it to be DAS in response to *Pst* and Kumar & Kirti (2012) observed that out of SA ABA and methyl jasmonate only methyl

jasmonate resulted in a significant and consistent induction of *RdRSZ21,* with transcripts accumulating in a time dependant manner.

Ding *et al* (2014) carried out genome-wide RNA-Seq analysis on Arabidopsis response to NaCl where they found that 10 of the 14 SR genes undergo DAS, but none showed DGE. This fits with our data that indicates that the SR proteins are predominantly regulated by DAS in response to stress. However, Howard *et al* (2013) carried out genome-wide RNA-Seq analysis on Arabidopsis infected with *Pst* and only found one SR gene to undergo DAS (SR30), with two SR genes (SR30 and SR34b) that were DE. This study identified DAS by only utilising isoform level analysis. Looking at our data only two SR proteins were detected at the isoform level (the majority were detected at the event level), once again indicating that more than isoform level analysis is required to adequately identify DAS genes.

4.4.10 Genes associated with chitin perception undergo *B. cinerea*-mediated DAS

The gene that encodes one of the most well studied chitin receptors, CERK1, was identified as undergoing *B. cinerea*-mediated DAS. The main proportional change identified was between the functional isoform *CERK1.1*, which increased in infected samples, and an isoform predicted to be targeted for NMD, *CERK_ID4*, which decreased in infected samples. As mentioned in section 4.4.5202 DAS coupled to NMD could enable a rapid response to infection by circumventing the time required for transcriptional activation and pre-mRNA accumulation. Thus we propose that, in response to infection, *CERK1* undergoes DAS thereby increasing the amount of the functional isoform, resulting in a rapid response of protein increase. In turn this increases the number of receptors that could detect the PAMP chitin, which could increase the likelihood of reaching the threshold to activate defence-related genes, although experimental evidence would be required to confirm this.

Two genes have been shown to be required for the proper splicing of *CERK1*, *SUA* and *RSN2*. Interestingly *SUA* is also shown to undergo *B. cinerea*-mediated DAS in our dataset and RSN2 was detected to undergo DAS in the Cooke (2012) dataset, potentially indicating it is DASed at a later stage of infection. In addition, one of the motifs that we detected as enriched in *B. cinerea*-mediated RI events, AATYATCAAG,

contains a motif, TCATC, which has been shown to bind to human RBM5, the plant homologue of which is *SUA*. This indicates that *B. cinerea*-mediated DAS forms part of the known complex network of defence-related genes.

In conclusion, infecting Arabidopsis leaves with *B. cinerea* results in genome-wide DAS, that is highly likely to have a functional role, due to key *B. cinerea* defence-related gene being affected. DAS that alters protein levels by either being coupled to NMD or by altering the 5'UTR could result in a rapid defence response because it would circumvent the time required for transcriptional activation and pre-mRNA accumulation. An investigation into the regulatory mechanisms of *B. cinerea*-mediated DAS has identified two candidate SR proteins that could be specific to biotic stress, although further work would be required to investigate this. In addition, a variety of putative SREs have been identified. Again further would is required to determine if they do indeed bind RBP.

5 Chapter 5: The spliceosome and the defence response; investigating the MOS4 associated complex.

5.1 Introduction

The spliceosome is a large ribonuclear protein complex whose structure consists of dynamically interacting snRNAs and hundreds of proteins, which work together to produce constitutive and alternative splicing (AS) (Reddy *et al*., 2012). An investigation into all of the spliceosomal and splicing associated factor genes is beyond the scope of this thesis, therefore we have focused on one spliceosome-associated complex, the MOS4 associated complex (MAC). The MAC is an evolutionary conserved spliceosome-associated nuclear complex homologous to the NineTeen Complex (NTC) in yeast and the PRP19 complex (also called the CDC5L complex) in humans (Monaghan *et al*., 2009). We chose this complex to investigate how spliceosome-associated complexes affect the defence response because there are a relatively small number of MAC components (24), it is involved in spliceosome assembly, and core components of the MAC are required for plant innate immunity as well as proper splicing of R genes (Monaghan *et al*., 2009; Palma *et al*., 2007; F. Xu *et al*.,2012).

Aims

The overarching aim of this chapter is to start to elucidate some of the mechanisms underlying defence responsive splicing, with the focus on an evolutionarily conserved spliceosome associated complex, the MAC. To achieve this the following research objectives will be met:

1. Perform morphological and defence phenotype screens on MAC loss-of-function mutants to
    a. Identify components of the MAC that are linked to *B. cinerea* defence, i.e. which MAC loss-of-function mutants have altered susceptibility of *B. cinerea* and
        i. Determine if this defence phenotype is linked to morphological changes in the loss-of-function mutants
        ii. Determine if the defence phenotype is specific to *B. cinerea* or occurs against other pathogens
2. Determine if loss-of-function MAC mutants affect *B. cinerea* DAS by
    a. Ascertaining if loss-of-function MAC mutants with a defence phenotype alter the splicing of *AT4G39270*
    b. Identifying two MAC loss-of-function mutants to investigate at the genome-wide level to
        i. Determine if infection-responsive DAS depends upon specific subunits of the MAC
        ii. Investigate if different MAC components play divergent roles in infection specific splicing, i.e. are subsets of DAS events affected in different MAC mutants

4. Ascertain if MAC components influence other splicing factor associated genes.

5.2 Results

In Chapter 4 we showed that genes encoding components of the MAC are differentially expressed and undergo DAS in response to *B. cinerea* 24hpi (Table 24).

Table 24. Components of the MAC can undergo DGE and are DASed in response to *B. cinerea* 24hpi. Four MAC component genes are DE in response to *B. cinerea* (indicated by tick in column four), and thirteen are predicted to undergo DAS (indicated by tick in column five).

Gene	Arabidopsis Gene Name	Human Homologue Gene Name	DEG	DAS
AT1G04510	Mos4-Associated Complex 3a (**MAC3A)**	Pre-MRNA Processing Factor 19 (**PRP19**)		✓
AT1G06220	Maternal Effect Embryo Arrest 5 (**MEE5**)	Elongation Factor Tu Gtp Binding Domain Containing 2 (**EFTUD2**)		
AT1G07360	Mos4-Associated Complex 5a (**MAC5A)**			
AT1G09770	ARABIDOPSIS THALIANA CELL DIVISION CYCLE 5 (**ATCDC5)**	CELL DIVISION CYCLE 5 (**CDC5)**		✓
AT1G10580		Cell Division Cycle 40 (**CDC40**)	✓	✓
AT1G15200				✓
AT1G20960	Embryo Defective 1507 (**EMB1507**)	Small Nuclear Ribonucleoprotein 200kda (**SNRNP200**)		✓
AT1G34290	Receptor Like Protein 5 (**RLP5**)			
AT1G77180	Snw/Ski-Interacting Protein (**SKIP**)			✓
AT1G80070				✓
AT2G33340	Mos4-Associated Complex 3b (**MAC3B)**	Pre-Mrna Processing Factor 19 (**PRP19**)		✓
AT2G38770	Embryo Defective 2765 (**EMB2765**)	Aquarius Intron-Binding Spliceosomal Factor (**AQR**)		
AT2G43770		Small Nuclear Ribonucleoprotein 40kda (U5) (**SNRNP40**)		
AT3G15730	Phospholipase D Alpha 1 (**PLDALPHA1**)			✓
AT3G18165	Modifier Of Snc1,4 (**MOS4)**	**Pre-Mrna-Splicing Factor** Spf27 Or Breast Carcinoma Amplified Sequence 2 (**BACS2**)		
AT3G18790		ISY1-RAB43		
AT3G20820			✓	
AT3G60190	Arabidopsis Dynamin-Like 4			✓
AT4G15900	Pleiotropic Regulatory Locus 1 (**PRL1)**	Pleiotropic Regulator 1 (**PRLG1**)		
AT4G19410	Pectin Acetylesterase 7	Notum	✓	✓
AT5G28740		XPA Binding Protein 2 (**XAB2**)	✓	
AT5G41770		Crooked Neck Pre-Mrna Splicing Factor 1 (**CRNKL1**)		
AT5G42080	Dynamin-Like Protein			✓
AT5G64270		Splicing Factor 3b, Subunit 1, (**SFB1**)		✓

To investigate the individual components of the MAC, seeds were obtained of mutant knock-out (KO) lines for each of the 24 components. Four homozygote lines were obtained from published lines (*Atcdc5*$_{KO}$, *mos4*$_{KO}$ and *prl1*$_{KO}$ (Monaghan *et al*., 2009), *mac5a*$_{KO}$ (Monaghan *et al.*, 2010)). The remaining twenty lines were obtained from Nottingham Arabidopsis Stock Centre. All 24 lines were screened to identify homozygous mutants (Figure 74). Of these 24 lines, four lines produced no homozygous seedlings and were thus deemed to be homozygote lethal (*AT1G20960*$_{KO}$, *AT1G77180*$_{KO}$, *AT1G80070*$_{KO}$ and *AT5G42080*$_{KO}$). Seeds were harvested from the twenty confirmed homozygote KO lines and used to grow plants for all further experiments.

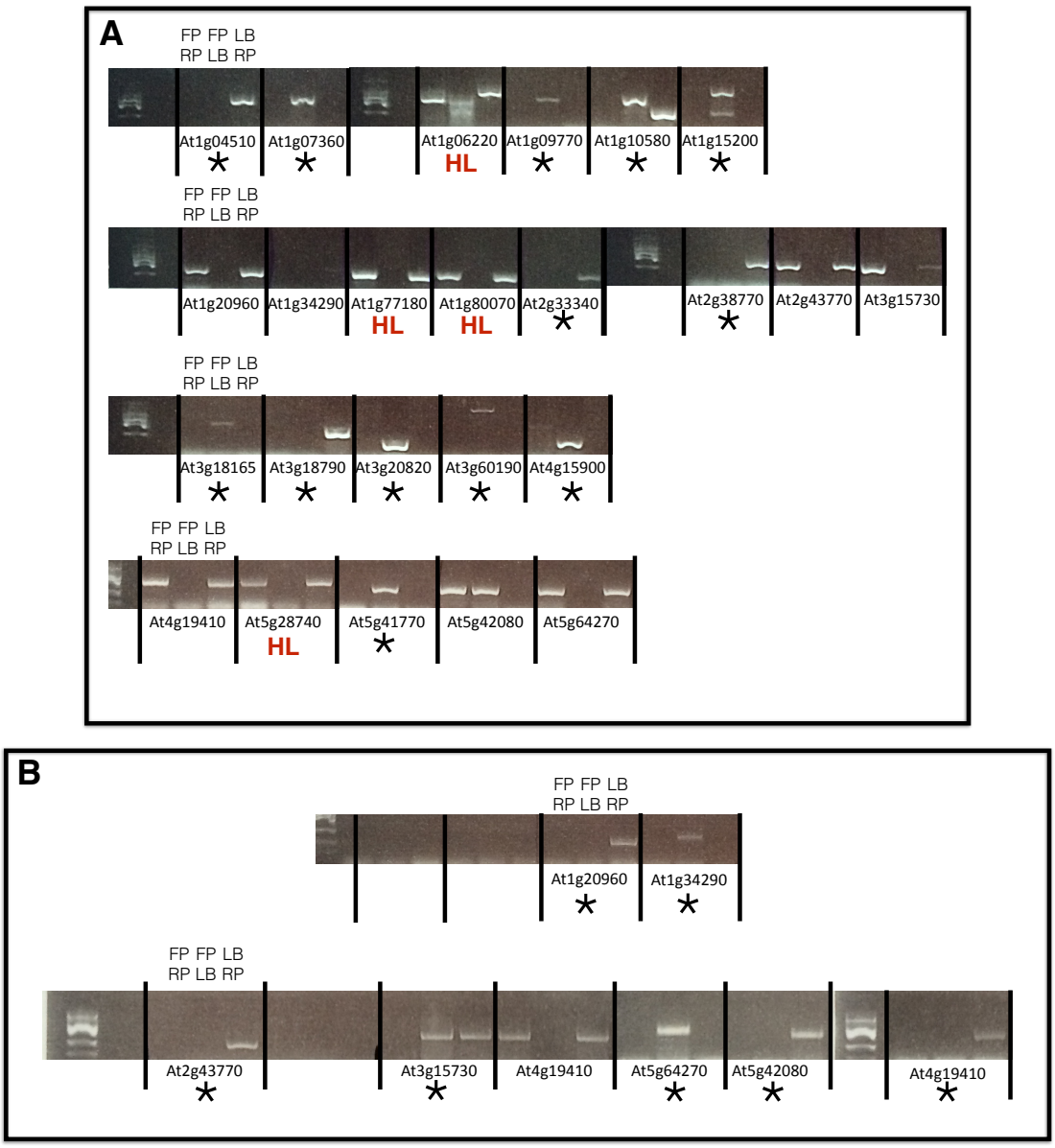


Figure 74. Homozygous mutants are identified for all bar four components of the MAC. A. Batches of ten seedlings were tested for homozygosity five times per line, thirteen lines tested positive for homozygotes indicated by *. B. For those lines where no batches where homozygous, individual plants were tested at three weeks. Four lines were deemed to be putative homozygote lethal due to experiments on five batches and ten individual plants not detecting any homozygotes (depicted HL in A). Black lines divide samples; each sample is run with three primer pairs with columns 1 to 3 being: 1, forward primer (FP) and reverse primer (RP) to detect wild-type, 2, FP and left border of T-DNA (LB) to detect presence of T-DNA insertion, 3, RP and LB to detect presence of T-DNA insertion. Please not this is a composite image; individual gel photographs are available on request.

5.2.1 Six individual MAC knock-out mutant plants have increased susceptibility to *B. cinerea*.

The 20 homozygote KO lines were screened for increased susceptibility to *B. cinerea* using leaf 7 of four-week–old plants to select potential candidates involved in the mechanisms of DAS in response to *B. cinerea* infection. Of these 20 homozygote KO lines it was noted that one line $At1g07360_{KO}$ flowered early and 10 lines had a reduced number of leaves at four weeks (Figure 75). This indicates that mutations of MAC components affect plant morphology.

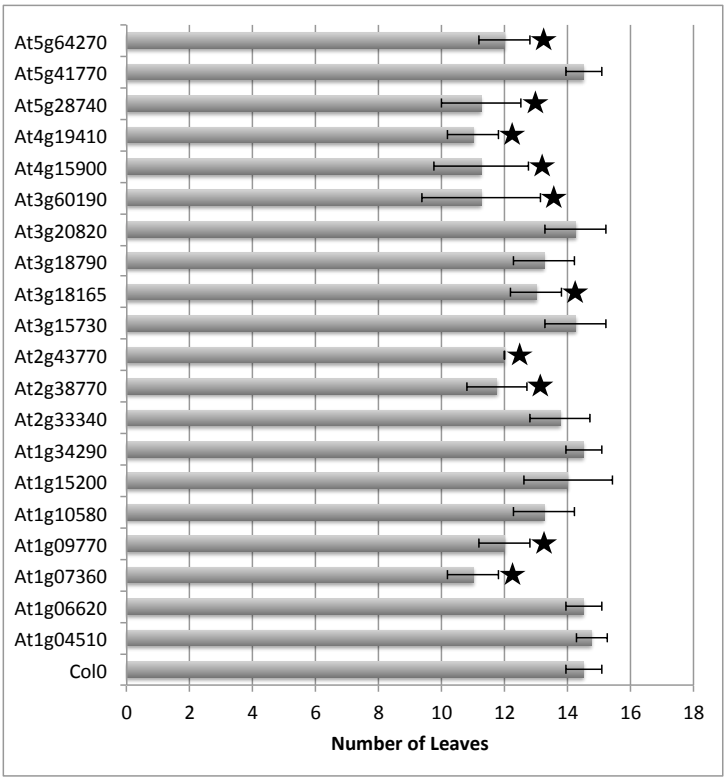


Figure 75. Ten MAC mutants had reduced number of leaves in four-week-old Arabidopsis plants. Stars indicate reduced number of leaves compared to wild-type (Students t-test, $p<0.05$); error bars represent standard deviation of four biological replicates.

Six KO lines had increased susceptibility to *B. cinerea* compared to wild-type plants at 48 and 72 hpi, $At1g07360_{KO}$ ($mac5a_{KO}$), $At1g09770_{KO}$ ($Atcdc5_{KO}$), $At2g38770_{KO}$, $At3g18165_{KO}$ ($mos4_{KO}$), $At3g20820_{KO}$ and $At4g15900_{KO}$ ($prl1_{KO}$) (Figure 76). This indicates that components of the MAC are involved in plant defence against *B. cinerea*. Mutations in the previously characterized MAC genes *MOS4*, *AtCDC5*, *PRL1*

and the double mutants *mac3a mac3b* result in enhanced disease susceptibility to *Hyaloperonospora arabidopsidis* (*Hpa*) *Noco2*, an obligate biotrophic oomycete, and two strains of the hemibiotrophic bacterial species, *Pseudomonas syringae pv maculicola* (*Psm*) and *pv* tomato (*Pst*) DC3000 (Monaghan *et al*., 2009; Palma *et al*., 2007), but to our knowledge this is the first time they have been implicated in defence of a necrotrophic organism.

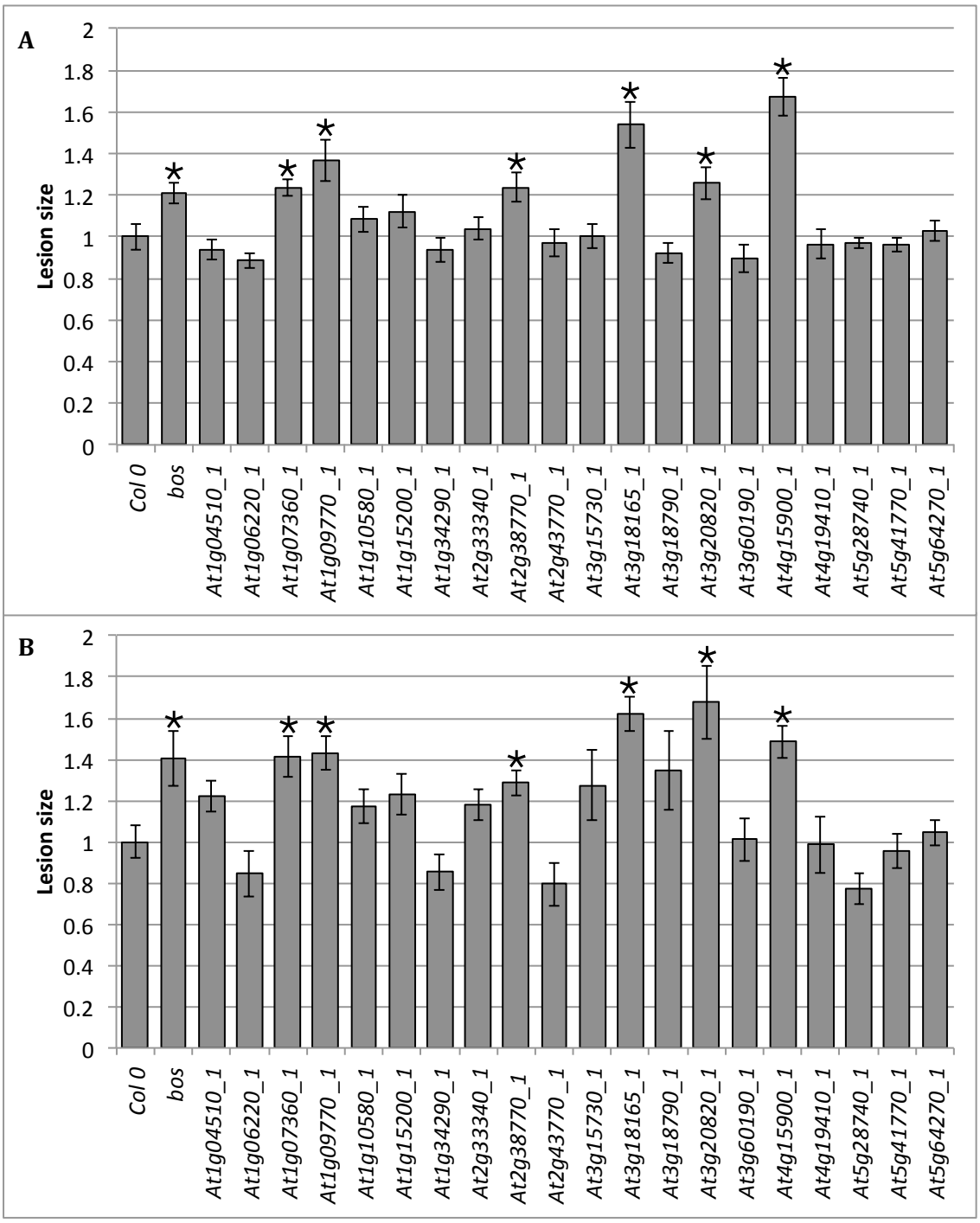


Figure 76. ***B. cinerea*** **phenotype screen identified six MAC KO lines that displayed a phenotype. Six lines,** ***At1g07360*** $_{KO}$ **(*mac5a*** $_{KO}$**),** ***At1g09770*** $_{KO}$ **(*Atcdc5*** $_{KO}$**),** ***At2g38770*** $_{KO}$**,** ***At3g18165*** $_{KO}$ **(*mos4*** $_{KO}$**),** ***At3g20820*** $_{KO}$ **and** ***At4g15900*** $_{KO}$ **(*prl1*** $_{KO}$**) are statistically more susceptible to** ***B. cinerea*** **than the Col0 wild-type at 48 hpi (A) and 72 hpi (B) indicated by the star (Students t-test p<0.05, biological replicates = 27). Error bars indicate SEM.** ***bos1*** **was used as the positive control. Lesions size are normalised to Col0 wild-type to enable direct comparisons.**

For the six MAC KO lines that had increased susceptibility to *B. cinerea* RT-qPCR was used to determine the percentage knockout of the gene. All six lines had a significant reduction or complete knockout of the gene (Figure 77).

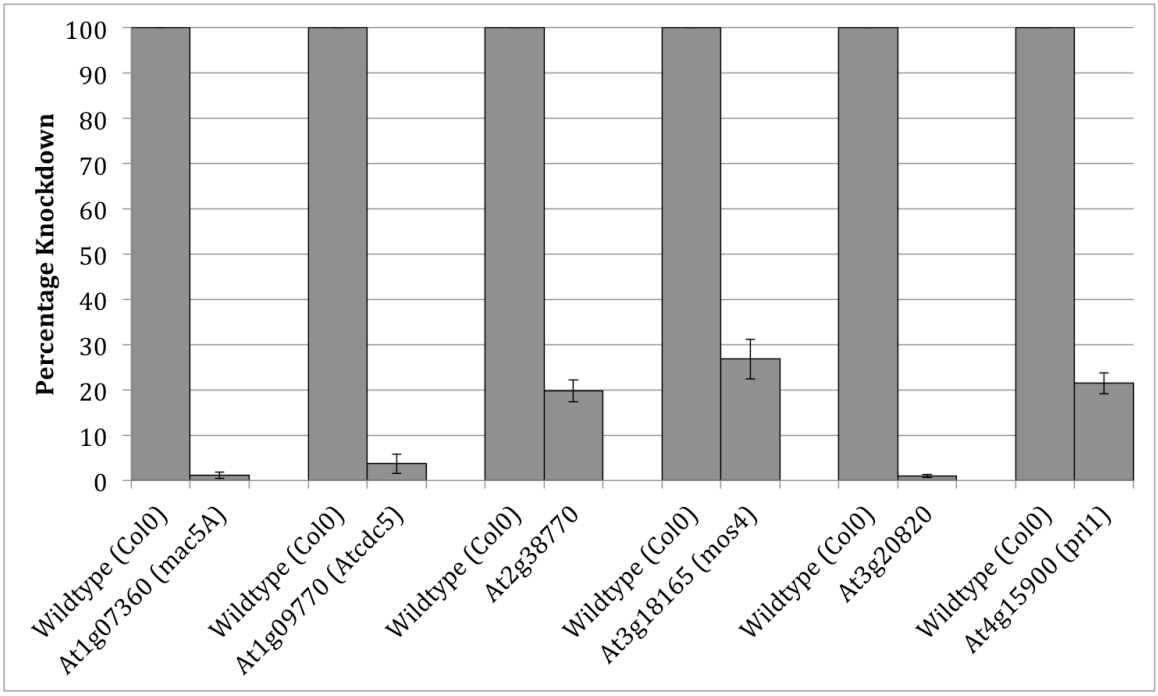


Figure 77. Percentage knockdown of homozygous KO lines that are more susceptible to *B. cinerea*. Each of the six lines is significantly reduced in gene expression of the KO gene compared to Col0 Arabidopsis wild-type plants (Students t-test p<0.05), with *mac5a*$_{KO}$, *Atcdc5*$_{KO}$ and *At3g20820*$_{KO}$ being considered as complete KO. Error bars are SEM of four biological replicates of leaf 8 from four-week-old plants. Gene expression was normalised to *PUX1 (AT3G27310)* a gene known not to change in response to *B. cinerea* infection (Windram *et al.*, 2012).

Our screen of the MAC KO T-DNA mutants has identified six potential genes involved in the defence response against *B. cinerea* as demonstrated by their increased susceptibility, and we have shown that these T-DNA KO lines have significantly reduced expression of the gene of interest. However T-DNA transformants contain an average of 1.4 inserts each (Feldmann, 1991). Thus to ensure that our genes of interest are the one producing the *B. cinerea* phenotype for non-backcrossed lines (*At1g07360*, *At2g38770* and *At3g20820*), additional KO lines were obtained and screened for homozygosity (Figure 78). This was not required for the backcrossed lines because backcrossing should remove any potential unlinked, additional T-DNA or genomic disruptions. All additional lines of *At2g38770* were homozygote lethal,

this agrees with Tzafrir *et al* 2003 who found mutants of this gene to be lethal (Tzafrir *et al*., 2003), the homozygous line that we identified is only knocked down by 80%, potentially indicating that this small amount of gene expression is enough to overcome the lethal mutation.

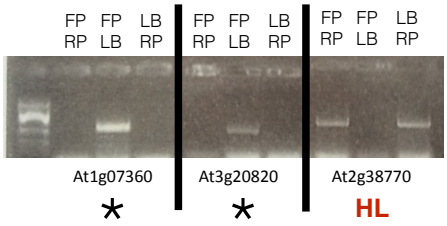


Figure 78. Homozygosity check for second line of *At1g07360*, *At2g38770* and *At3g20820* mutants. The second line for the At1g07360 and At2g38770 were homozygote. However all, plants tested for At2g38770 were heterozygote, indicating that it is homozygote lethal (one example is given here). Black lines divide samples; each sample is run with three primer pairs with columns 1 to 3 being: 1, forward primer (FP) and reverse primer (RP) to detect wild-type, 2, FP and left border of T-DNA (LB) to detect presence of T-DNA insertion, 3, RP and LB to detect presence of T-DNA insertion. Please note this is a composite image; individual gel photographs are available on request

A finer time course *B. cinerea* phenotype screen was run with an additional KO line added for those MAC genes that were not from back crossed lines (*At1g07360* and *At3g20820*). Because the second line of *At2g38770* was homozygote lethal a second line could not be tested. At 34 hpi only the positive control *bos1* and the *prl1*$_{KO}$ mutants were statistically more susceptible to *B. cinerea* than the Col0 wild-type leaves (T-test p=0.049 and p=0.041 respectively) but lesions are still small at this stage. By 48 hpi all of the MAC KO lines tested were statistically more susceptible to *B. cinerea* than the Col0 wild-type, as was *bos1* ($p<0.01$ for all lines), this was also the case at 72 hpi (T-test $p<0.001$ for all lines) (Figure 79). This shows that these six components of the MAC play a role in the plant defence against the necrotrophic pathogen *B. cinerea*.

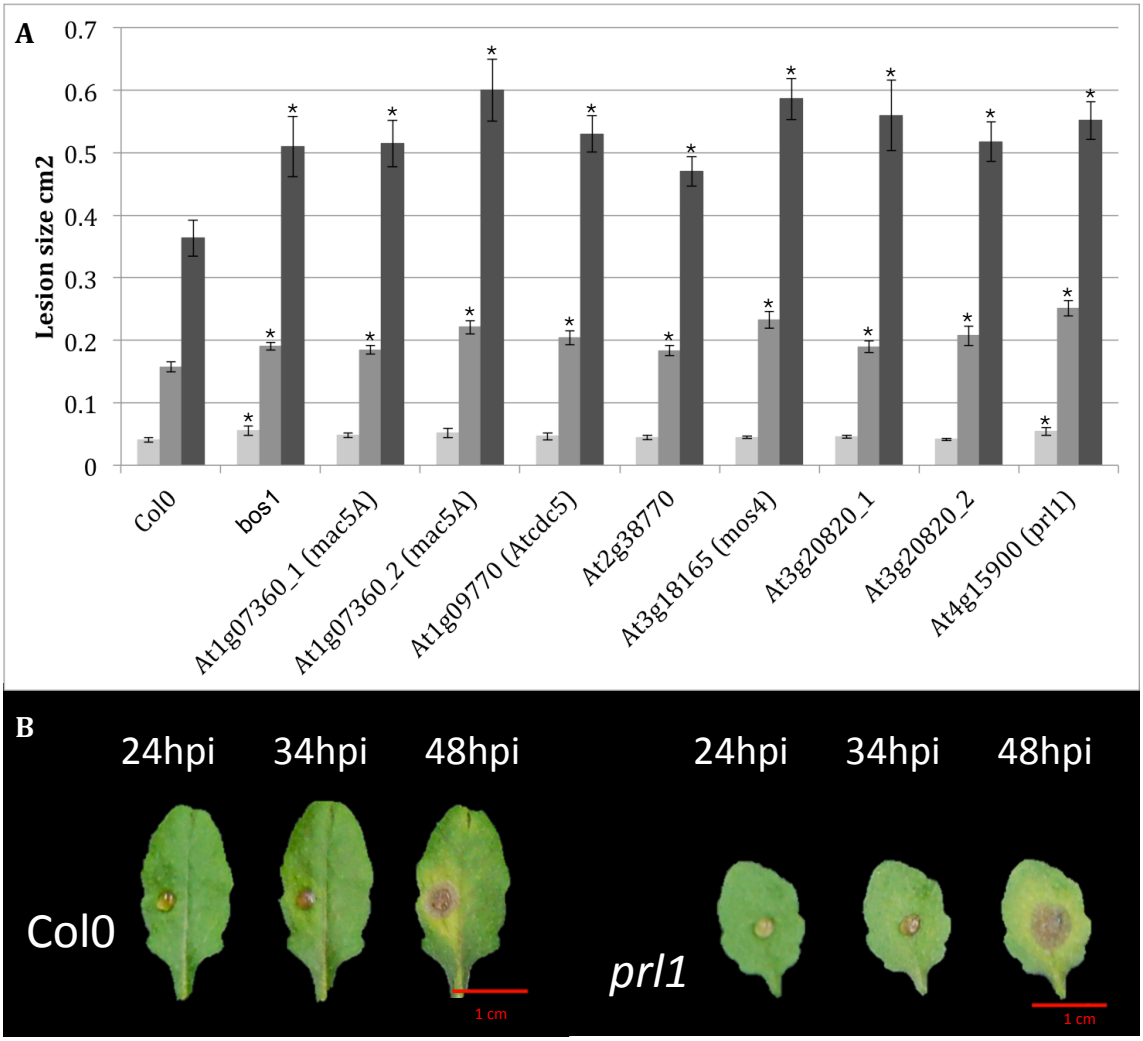


Figure 79. Six MAC genes when knocked out increase susceptiblity to *B. cinerea*. A. At 34 hpi only *prl1*$_{KO}$ and the positive control (*bos1*) are statistically more susceptible to *B. cinerea* 24 hpi compared to wild-type plants (Students t-test p=0.041 and p=0.049 respectively, number of leaves (n) = 25). By 48 hpi all the previously found MAC KO mutants that have increased susceptibility to *B. cinerea* plus the additional lines for the non-backcrossed MAC KO genes lines *mac5a_2* and *At3g20820_2* displaying increased susceptibility to *B. cinerea* compared to wild-type plants (Students t-test all lines p<0.01, n=25), this was also the case at 72 hpi (Students t-test all lines p<0.001, n=25). B. The MAC KO line *prl1*$_{KO}$ statistically larger lesion size compared to Col0 wild-type plants after 34hpi are hard to observe visually; by 48hpi the increased lesion size is clearly visible.

5.2.2 Some MAC loss-of-function transgenic plants that display increased susceptibility to *B. cinerea* have an altered morphology compared to wild-type plants.

During the *B. cinerea* susceptibility screen it was noted that some of mutant plants had altered morphology. To investigate this further, morphological measurements

were taken of plants grown under two different conditions; ten-day-old seedlings grown in agar, and in five-week-old plants grown in soil.

5.2.3 Three single MAC KO mutants with a *B. cinerea* phenotype weigh less as ten-day-old-seedlings and have altered root structure.

Four of the MAC KO mutants with increased susceptibility to *B. cinerea* had significantly shorter roots than wild-type Col0 seedlings, $mac5a_{KO}$, $Atcdc5_{KO}$, $mos4_{KO}$ and $prl1_{KO}$. Of these four, three also had a decreased dry weight compared to wild-type Col0 plants, $Atcdc5_{KO}$, $mos4_{KO}$ and $prl1_{KO}$, at the 5% level (Figure 80).

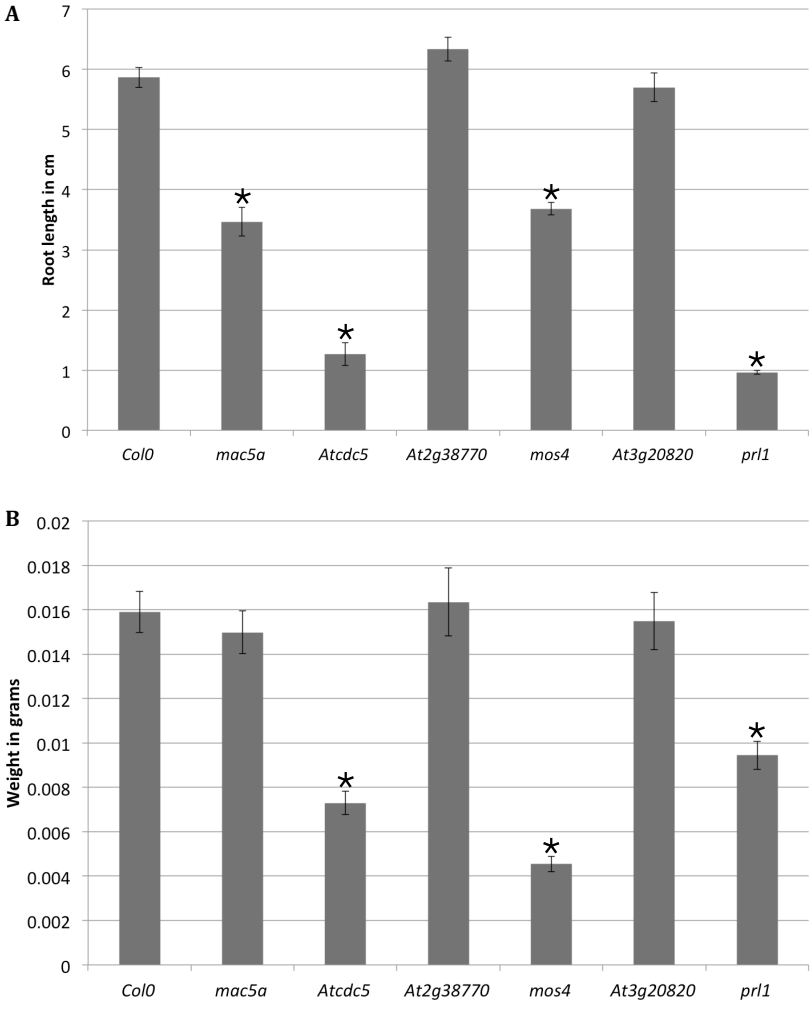


Figure 80. Four MAC knockout mutants have decreased root length with three having reduced dry weight in ten-day-old-seedlings. A. Compared to wild-type Col0 seedlings the MAC KO mutants $mac5a_{KO}$, $Atcdc5_{KO}$, $mos4_{KO}$, $prl1_{KO}$ have shorter roots, with $prl1_{KO}$ having the shortest (Students t-test p<0.05, number of biological replicates (n)=20). B. Three of these also have decreased dry weight compared to wild-type Col0 plants ($Atcdc5_{KO}$, $mos4_{KO}$, and $prl1_{KO}$), with $mos4_{KO}$ being the lightest (Students t-test p<0.05, n=12). Error bars are SEM, stars indicate statistically significantly different from wild-type (Col0).

5.2.4 Four five-week-old soil grown MAC KO mutants with a *B. cinerea* phenotype have altered morphology compared to wild-type Col0 plants.

Soil grown five-week-old plants were chosen to investigate the morphological changes, because this was the time when the flower buds had just set in the wild-type Col0 plants. In the MAC KO lines $mac5a_{KO}$ was already flowering, whereas $Atcdc5_{KO}$ and $mos4_{KO}$ did not have flower buds set (Figure 81A). This corresponded to them flowering later than the other MAC KO lines and Col0. The $mac5a_{KO}$ had slightly elongated leaves that often appeared twisted, whereas the $Atcdc5_{KO}$, $mos4_{KO}$ and $prl1_{KO}$ had smaller leaves than wild-type plants with $Atcdc5_{KO}$ and $prl1_{KO}$ having slightly crinkly edges and $mos4_{KO}$ having rounded leaves, which agrees with their published morphological descriptions (Monaghan *et al*., 2010; Palma *et al*., 2007). $AT2G38770_{KO}$ and $AT3G20820_{KO}$ had similar morphology to wild-type plants (Figure 81A). The $Atcdc5_{KO}$, $mos4_{KO}$ and $prl1_{KO}$ lines were observed to be smaller than the wild-type. To investigate this, images were taken and the area covered by the plant measured using the ImageJ software; the $mac5a_{KO}$, $Atcdc5_{KO}$, $mos4_{KO}$ and $prl1_{KO}$ lines were all significantly smaller than Col0 wild-type plants (Figure 81B). The additional KO lines for *MAC5A* and *AT3G20820* showed the same morphology as the previous $mac5a_{KO}$ and the $At3g20820_{KO}$ lines (Figure 82). Taken together this work indicates that some of the components of the MAC have morphological changes as well as increased susceptibility to *B. cinerea,* indicating that unsurprisingly the MAC has a broader role than that of defence.

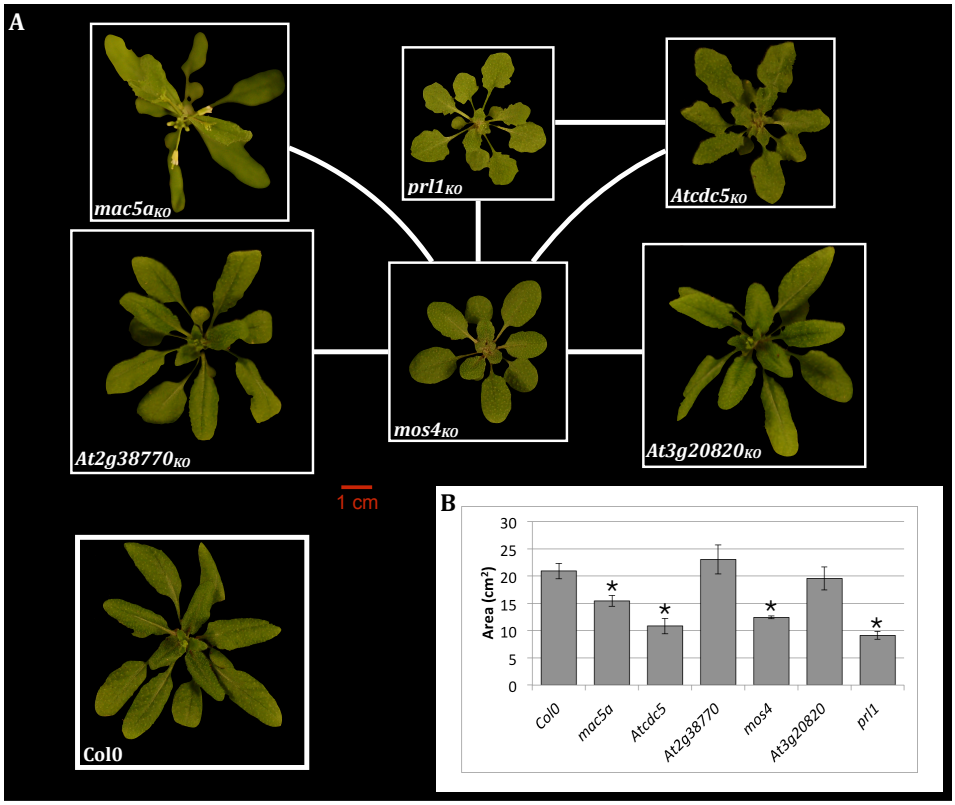


Figure 81. Morphological comparisons of five-week-old *B. cinerea* susceptible MAC KO mutants to wild-type Col0 plants. A. Photograph of five-week-old soil grown plants. Size bar represents 1 cm. The size of the rectangle surrounding the plant is proportional to the mean plant area of six biological replicates. The interconnecting lines indicate how they interact *in planta*. B. The KO lines *mac5a*$_{KO}$, *Atcdc5*$_{KO}$, *mos4*$_{KO}$ and *prl1*$_{KO}$ cover a statistically significantly smaller area than wild-type Col0 plants. Values represent the mean area of six biological samples ±SEM. Students t-test was used to analyse the statistical significance of area size compared to Col0, Asterisk indicates p<0.01.

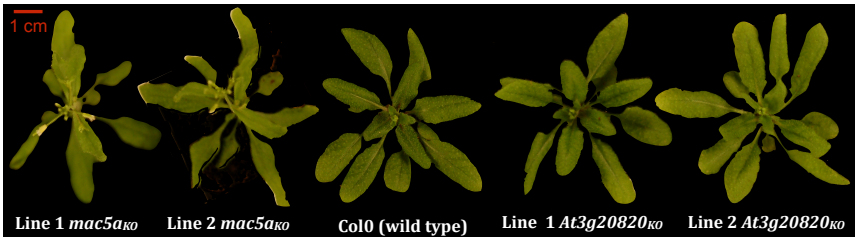


Figure 82. Photograph of five-week-old soil grown plants of two independent KO lines *mac5a* and *AT3G20820.* Size bar represents 1 cm. Both *mac5a* KO lines display altered morphology from wild-type plants with earlier flowering, whilst both At3g20820 KO lines have similar morphology to the Col0 wild-type plants.

5.2.5 *AT2G38770*, *MOS4*, and *PRL1* are involved in a range of pathogen defence responses

Previously Palma *et al* (2007) found that the MAC KO mutants *Atcdc5*$_{KO}$, *mos4*$_{KO}$ and *prl1*$_{KO}$ display increased susceptibility to *Hyaloperonospora arabidopsidis* (*Hpa*) *Noco2*, an obligate biotrophic oomycete, and two strains of the hemibiotrophic bacterial species, *Pseudomonas syringae pv maculicola* (*Psm*) and *pv* tomato (*Pst*) DC3000, indicating that these MAC components are all required for basal defence (Palma *et al*., 2007). This indicates that *Atcdc5*$_{KO}$, *mos4*$_{KO}$ and *prl1*$_{KO}$ are involved in the plant defence response to a broad range of pathogens. To investigate the breadth of the pathogen defence of the other MAC mutants that we detected to have increased susceptibility to *B. cinerea* we examined them for susceptibility to *Pst* DC3000.

The MAC KO mutants *Atcdc5*$_{KO}$, *mos4*$_{KO}$ and *prl1*$_{KO}$ display increased susceptibility to *Pst* at 1 day post infection (dpi), 2 dpi and 3 dpi (as expected), whereas *At2g38770*$_{KO}$ only showed increased susceptibility at 3 dpi; *mac5a*$_{KO}$, and *At3g20820*$_{KO}$ did not show an increased susceptibility to *Pst* (Figure 83).

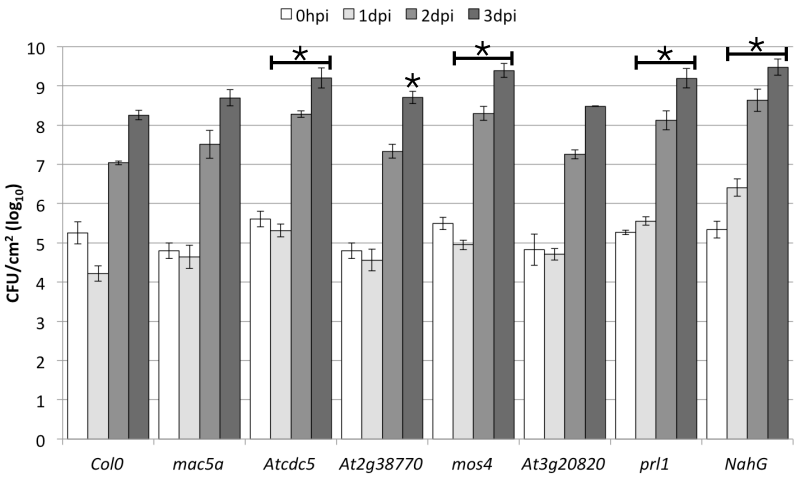


Figure 83. The Loss of *AtCDC5*, *AT2G38770*, *MOS4* and *PRL1* result in enhanced susceptibility to *Pst* as shown by bacterial growth at 0, 1 and 3 days post-inoculation. Values represent the average of six biological replicates ±SEM. Students t-test was used to analyse the statistical significance of bacterial growth compared to Col0, Asterisk indicates p<0.05. *NahG* was used as a positive control.

As expected we found that our positive control *NahG* and the three MAC KO mutants previously detected by Palma *et al* (2007) to have increased susceptibility to

Pst; *Atcdc5$_{KO}$*, *mos4$_{KO}$* and *prl1$_{KO}$*, were more susceptible to *Pst* infection at 1, 2 and 3 dpi. *At2g38770$_{KO}$* was also detected to have increased susceptibility to *Pst*, but due to it not being back-crossed and no additional homozygote independent T-DNA line being available for this gene, we cannot be certain that this phenotype is due to our gene of interest. The remaining two MAC mutants that had increased susceptibility to *B. cinerea*, *mac5a$_{KO}$* and *At3g20820$_{KO}$* did not display increased susceptibility to *Pst*. The work on *mac5a$_{KO}$* agrees with published data by Monaghan *et al* (2010), who state that the single mutant *mac5a$_{KO}$* does not exhibit increased susceptibility to *Pst*, probably due to partial redundancy with *MAC5B*. This potentially indicates that although this partial redundancy is enough to prevent increased susceptibility to *Pst*, it is not enough to protect against increased susceptibility to *B. cinerea*. This work indicates that the core components of the MAC, AtCDC5, MOS4, and PRL1 have a broad role in the plant defence response, whereas MAC5A and AT3G20820 may have a more specific role, potentially targeting fungal or necrotrophic pathogens, although further work would be required to confirm this.

5.2.6 *Atcdc5$_{KO}$* abolishes the DAS of the LRR-RLK *AT4G39270*

Having seen several components of the MAC impact on defence, we now want to examine if this occurs via differential alternative splicing (DAS) in response to *B. cinerea* infection. We have previously shown in Chapters three and four that in response to *B. cinerea* 24 hpi the proportion of the spliced intron (SI) isoform compared to the retained intron (RI) isoform of *AT4G39270*, a LRR-RLK, increases. To help determine if components of the MAC play a role in DAS in response to *B. cinerea* we investigated if those MAC components with increased susceptibility to *B. cinerea* had an effect on the relative proportions of the SI isoform relative to the RI isoform of intron one of *AT4G39270*. The seventh leaf of four-week-old plants were inoculated with mock and *B. cinerea* spore suspension, leaves were harvested 24 hpi and RT-PCR using isoform specific primers against SI and RI was performed. In mock samples the *prl$_{KO}$* had a higher relative quantity of the SI relative to the RI isoform compared to Col0 wild-type plants (p=0.005). However this did not affect its ability to increase the proportion of the SI relative to the RI isoform in response to *B.*

cinerea 24 hpi. All other lines did not have statistically different SI relative to RI levels compared to Col0 plants in mock samples. In the *Atcdc5$_{KO}$* the increase of the SI relative to the RI isoform in response to infection is abolished; there is no statistically significant difference between mock and *B. cinerea* 24hpi samples (Student t test p=0.99), resulting in significantly lower amounts of the SI relative to the RI in *Atcdc5$_{KO}$* infected samples compared to the Col0 wild-type infected leaves (Student t test p=0.03) (Figure 84). Although *mos4$_{KO}$* and *At3g20820$_{KO}$* appear to increase the relative amount of the SI compared to the RI this is not significant at the 5% level (Student t test p = 0.1 and 0.06 respectively). These results indicate that AtCDC5 is required for the infection-induced splicing of the LRR-RLK encoding gene *AT4G39270*.

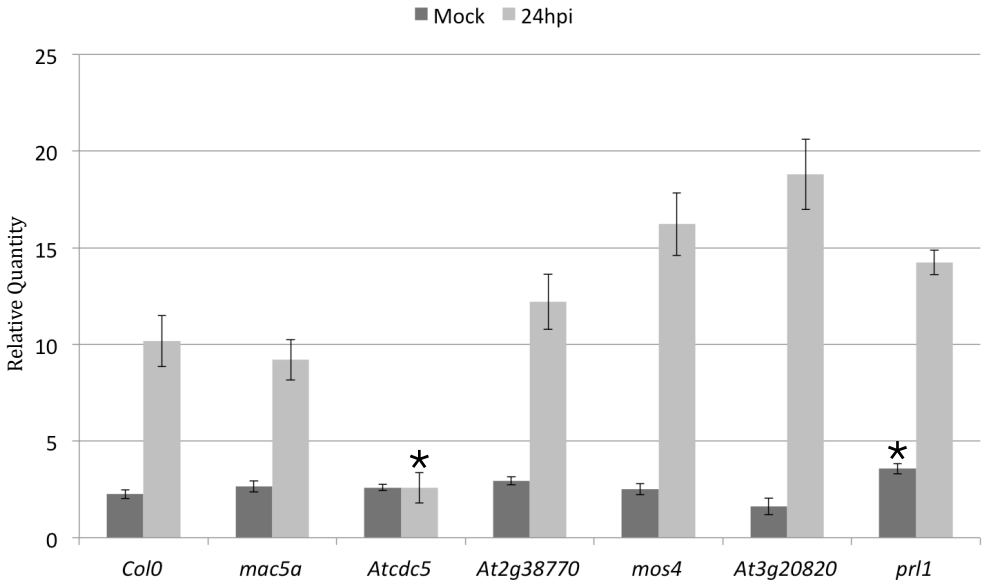


Figure 84. The quantity of the SI isoform relative to the RI isoform in MAC KO mutants. *Atcdc5$_{KO}$* has altered relative quantities in infected samples and *prl1$_{KO}$* has altered relative quantities in mock samples. Values represent an average of five biological replicates±SEM. Students t-test was used to analyse the statistical significance of the quantity of the SI isoform relative to the RI isoform compared to Col0, Asterisk indicates p<0.05.

5.2.7 Genome-wide analysis of *Atcdc5$_{KO}$* and *mos4$_{KO}$*

To investigate if the abolition of *B. cinerea*-mediated DAS is a genome-wide phenomenon of *Atcdc5$_{KO,}$* an RNA-Seq experiment was performed, comparing *Atcdc5$_{KO,}$ mos4$_{KO}$* (included because *MOS4* encodes the central component of the MAC but does not affect DAS of *AT4G39270* when KO) and Col0 wild-type plants. The

amount of inoculum was administered proportionally based on average leaf size; five five-microliter droplets of a suspension of *B. cinerea* spores (1 X 10^5 spores mL^{-1}) or mock inoculum were placed on detached leaf 7 from 4-week-old Col0 Arabidopsis plants, *mos4*$_{KO}$ and *Atcdc5*$_{KO}$ leaves were inoculated with three and two five-microliter droplets respectively (Figure 85). This was done in order to achieve a consistent density of infection in the leaf tissue across the lines. Larger leaves require a larger number of drops for the same amount of cells to be infected. Thus altering the number of drops of an inoculum in a size dependant manner should result in approximately the same number of cells being infected in each leaf sample.

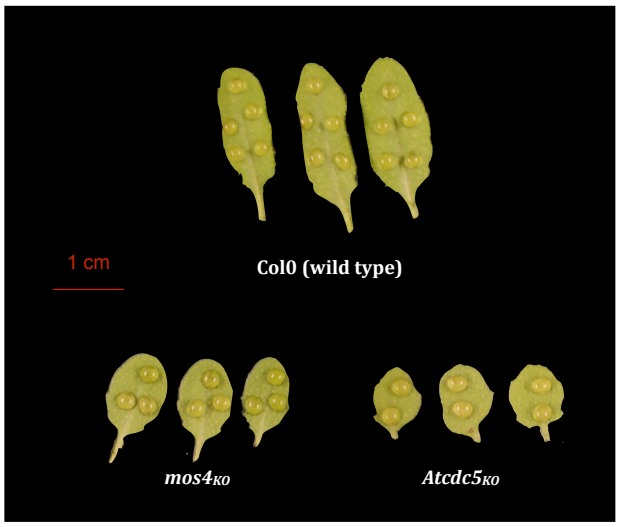


Figure 85. Photograph of infected plant material for RNA-Seq 24 hpi. Size bar represents 1cm.

Leaves were harvested at 24 hpi and RNA-Seq performed. Between 10-49 million sequenced paired end reads (100 bp in length) per sample were obtained using the Illumina High-Seq sequencing system (Table 25). The Col0 wild-type reads obtained from this data set have been investigated in Chapter 4 in order to investigate how *B. cinerea* affects the alternative splicing landscape of Arabidopsis. Now we compare that Col0 dataset with the *Atcdc5*$_{KO}$ and *mos4*$_{KO}$ data to begin to elucidate some regulatory mechanisms. The Col0, *Atcdc5*$_{KO}$ and *mos4*$_{KO}$ experiments were carried out at the same time with the same inoculum to minimise batch variation. Three biological replicates per treatment per lines were performed.

Table 25. Number of RNA-Seq reads. Between 10-49million paired end reads per sample were obtained.

Line	Treatment	Replicate	Number of paired end reads
Col0	Mock	1	26480741
Col0	Mock	2	29805886
Col0	Mock	3	28349890
Col0	B. cinerea 24hpi	1	25630664
Col0	B. cinerea 24hpi	2	49290487
Col0	B. cinerea 24hpi	3	31315650
$Atcdc5_{KO}$	Mock	1	16133816
$Atcdc5_{KO}$	Mock	2	14840872
$Atcdc5_{KO}$	Mock	3	19166379
$Atcdc5_{KO}$	B. cinerea 24hpi	1	19215578
$Atcdc5_{KO}$	B. cinerea 24hpi	2	20535056
$Atcdc5_{KO}$	B. cinerea 24hpi	3	23159410
$mos4_{KO}$	Mock	1	10213298
$mos4_{KO}$	Mock	2	12117997
$mos4_{KO}$	Mock	3	19527013
$mos4_{KO}$	B. cinerea 24hpi	1	19883640
$mos4_{KO}$	B. cinerea 24hpi	2	17311440
$mos4_{KO}$	B. cinerea 24hpi	3	19657370

Quality assessment with FastQC confirmed no processing or trimming of reads was required before RNA-Seq analysis commenced.

5.2.7.1 Knocking out Atcdc5 and mos4 affects differential gene expression in response to B. cinerea infection.

Although we are particularly interested in the splicing effects of MAC knockouts, we first investigated whether differential gene expression (DGE) per se was affected. The two pipelines (Tophat2->CuffDiff2 and STAR->CuffDiff2) utilised to detect DGE in Chapter 4 were used to detect DGE between mock and infected samples of each line independently. Comparison with *B. cinerea* marker genes compiled by Windram *et al* (2012) indicated that infection progressed as expected. Comparing genes DE between the three lines shows that approximately one quarter of Col0 genes that are DE in response to *B. cinerea* are affected by components of the MAC, i.e. they are DE in Col0 but not in either $Atcdc5_{KO}$ or $mos4_{KO}$ (Figure 86).

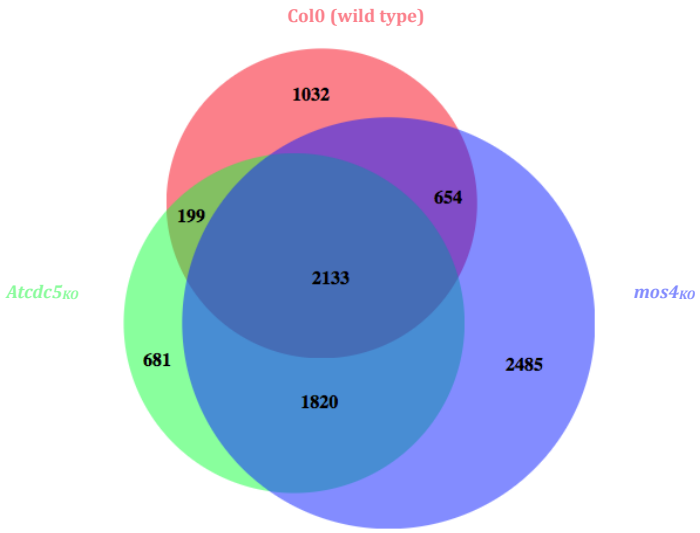


Figure 86. Venn diagram showing the number of genes DE (p<0.05) in response to *B. cinerea* 24 hpi in tested lines. 1032 (28% of Col0 DE genes) are not DE in either the *Atcdc5*$_{KO}$ or the *mos4*$_{KO}$, potentially indicating that these components of the MAC are involved in the regulation of these genes.

5.2.7.2 MAC loss-of-function mutants do not block whole scale AS or affect one particular event type

There is no significant difference between the total number of AS events regarding line or treatment (two way anova blocked by line and treatment p>0.05). RIs represent the largest proportion of individual events (between 26-29% of events), followed by A3SS (17-18%), then A5SS (9% in all cases) with SE having the smallest proportion of events (3-4%). There is no significant difference between the number of these events regarding line or treatment (two way ANOVA blocked by line and treatment p>0.05) (Figure 87). Although in the mock samples the percentage of RI events appears to be greater, this is not statistically significant indicating that the percentage and type of events is constant across the lines.

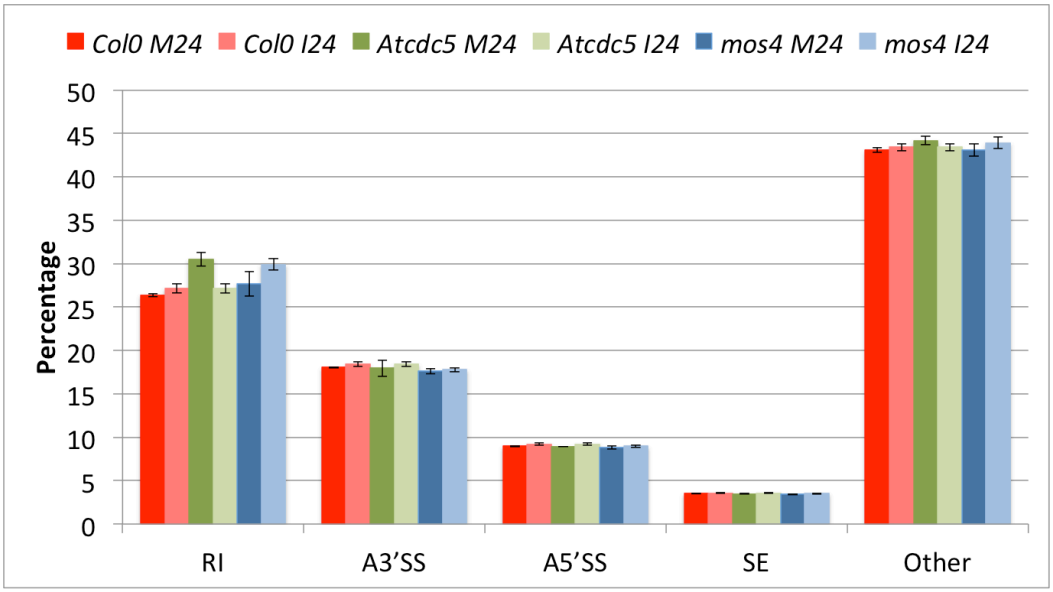


Figure 87. Percentage of each AS event type is not altered in loss-of-function *Atcdc5*$_{KO}$ and *mos4*$_{KO}$ mutants. All lines have approximately the same percentage of event types, with RI being the most prevalent event type; there is no statistical difference in the number of each event type between either lines or treatment (Two way anova blocked by treatment p>0.05). Dark blue, light blue, dark green, light green, dark red and light red represent Col0 mock, Col0 *B. cinerea* infected, *Atcdc5*$_{KO}$ mock, *Atcdc5*$_{KO}$ *B. cinerea* infected, *mos4*$_{KO}$ mock, and *mos4*$_{KO}$ *B. cinerea* infected samples respectively. Error bars represent the standard error of the mean of three biological samples.

These results indicate that there is the same amount of AS occurring in each of the lines. Thus, loss-of-function mutants *Atcdc5*$_{KO}$ or *mos4*$_{KO}$ do not completely abolish alternative splicing or block any one particular event type. However as seen above the DAS of *AT4G39270* is affected in *Atcdc5*$_{KO}$ lines, indicating that although the proportion of AS events detected are the same, loss of AtCDC5 may result in different isoforms being utilised in response to infection. This result was confirmed in the RNA-Seq data analysed using the event level methods, which showed that there was no significant change in the percentage spliced in (ΔPSI) of the first intron of *AT4G39270* in *Atcdc5*$_{KO}$ lines. The *mos4*$_{KO}$ mutant appears to increase the splicing of the first intron of *AT4G39270*, as it did in the RT-PCR experiment (Figure 88).

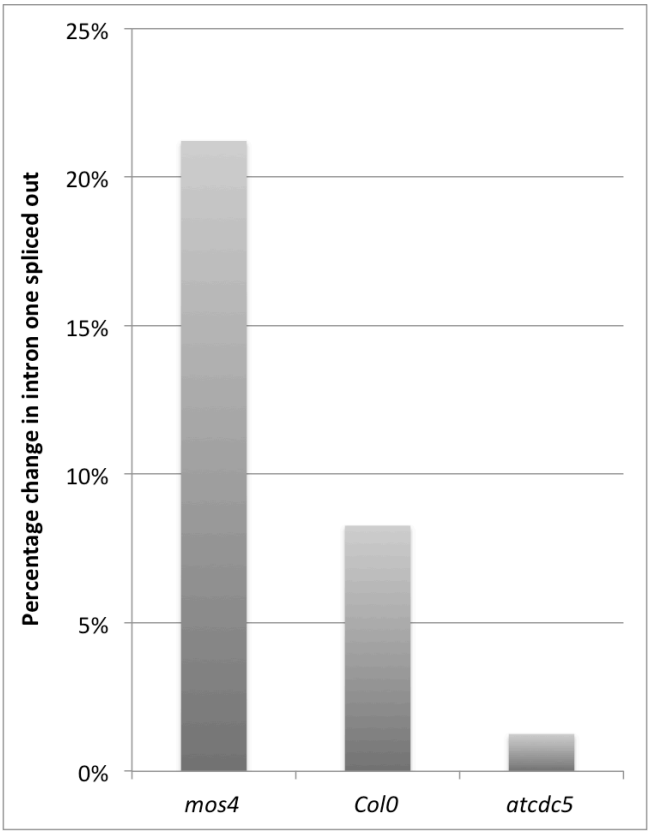


Figure 88. Percentage change of the first intron of AT4G39270 in response to *B. cinerea* infection. There is a higher percentage change in intron one spliced isoform, in the $Mos4_{KO}$ mutant compared to the wild-type Col0. The $Atcdc5_{KO}$ mutant has very little change in the percentage change of the intron one spliced isoform.

This is interesting because both $Atcdc5_{KO}$ and $mos4_{KO}$ have enhanced susceptibility to *B. cinerea*, but they affect the splicing of *AT4G39270* differently, indicating that the increased susceptibility seen in these mutants may be due to an accumulation of effects on *B. cinerea*-mediated DAS. To further investigate this the other validated RI events were examined to see if they displayed a similar trend.

5.2.7.3 $Atcdc5_{KO}$ and $mos4_{KO}$ affect the B. cinerea-mediated differential alternative splicing of validated defence-related genes

We have previously validated four additional RI events detected to undergo *B. cinerea*-mediated DAS by Cooke (2012). In chapter 4 we confirmed that these four RI events were DASed in our RNA-Seq dataset. Here we utilise these RI events to investigate the effect of the loss-of-function mutants, $Atcdc5_{KO}$ and $mos4_{KO}$ on the ΔPSI of RI events in response to *B. cinerea* infection. The majority of these validated DASed RI events are abolished in one or more of the loss-of-functions mutants, the

exception is *AT1G74590*, where all lines show a reduction in the RI event in response to *B. cinerea* infection (Figure 89).

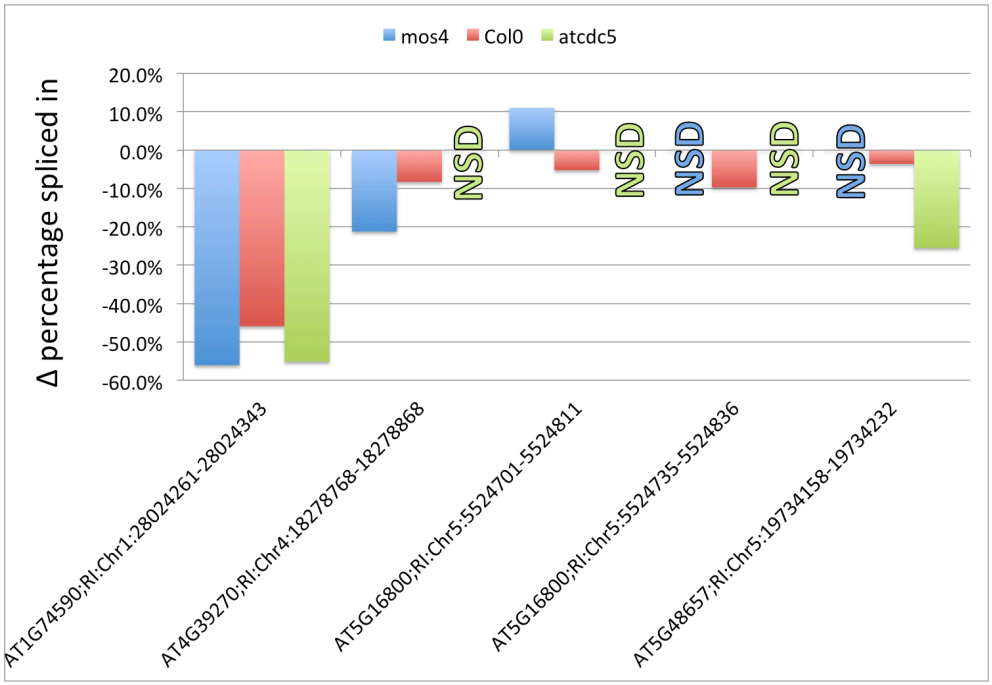


Figure 89. Both *Atcdc5$_{KO}$* and *mos4$_{KO}$* can affect DAS of defence-related genes. Values represent the ΔPSI in response to *B. cinerea* 24 hpi. Statistical significance was calculated using the SUPPA software, by comparing the observed ΔPSI between conditions with the distribution of the ΔPSI between replicates as a function of the gene expression. NSD indicates that there is no statistically significant difference in the percentage spliced in, i.e. there is no evidence that DAS occurs.

For all of the validated RI events, *B. cinerea* infection results in a decrease of the RI event, i.e. more of the spliced transcript is produced. *Atcdc5$_{KO}$* and *mos4$_{KO}$* affect both DAS RI events in *AT5G16800*, with both RI events being abolished in the *Atcdc5$_{KO}$*. However, in *mos4$_{KO}$* transgenic plants one of the RI events is abolished (Chr5:5524735-5524836) and the other RI event (Chr5:5524701-5524811) increases in response to infection, whereas in wild-type plants the intron is spliced out in response to infection in both cases (i.e. the ΔPSI will be negative). The remaining validated DASed RI event that occurs in gene AT5G48657 is abolished in *mos4$_{KO}$* transgenic plants, whereas in *Atcdc5$_{KO}$* transgenic line the RI event is reduced, as it is in the wild-type plants, although the ΔPSI is greater in the *Atcdc5$_{KO}$* mutants. This investigation of the previously validated RI events indicates that *B. cinerea*-mediated DAS is severely altered in *Atcdc5$_{KO}$* and *mos4$_{KO}$* transgenic plants suggesting that

AtCDC5 and MOS4 may play a functional role in regulating *B. cinerea*-mediated DAS. However, it does not show all *B. cinerea*-mediated DAS to be abolished. There appear to be specific effects depending upon the mutant and the gene. This is further evidence that there is no direct correlation between the effect on *B. cinerea*-mediated DAS and the *B. cinerea* defence phenotype. $Atcdc5_{KO}$ and $mos4_{KO}$ affect RI events differently but they both display increased susceptibility to *B. cinerea* infection, indicating that the defence phenotypes displayed by these mutants is an accumulation of overall effects on DAS.

Investigating the differences between the wild-type and two MAC KO lines may help us begin to understand how they contribute specifically to the regulation of splicing in response to *B. cinerea* infection.

5.2.7.4 *B. cinerea-mediated differential alternative splicing of Arabidopsis genes is altered in the MAC loss-of-function mutants at the genome-wide level*

To investigate the differences in DAS between wild-type and the two MAC loss-of-function mutants the event level analysis approach was utilised. This is because both the SJ and isoform level approach showed some ambiguity between DAS and DGE, and because we want to investigate the type of functions and gene families affected, rather than obtain an exhaustive list of genes, an event level analysis should suffice. Crucially this method identifies the change in *B. cinerea*-mediated DAS of *AT4G39270* in the MAC KO mutants seen in the RT-PCR experiment above (page 227).

At the genome-wide level $mos4_{KO}$ appears to increase the *B. cinerea*-mediated DAS, whereas in the $Atcdc5_{KO}$ decrease it. This also occurs at the individual event level (Figure 90). This is interesting considering that both mutants display an increased susceptibility to *B. cinerea* and other pathogens. Thus, it is plausible that it is the disruption of the balance of transcripts originating from a single gene that results in the defence phenotypes displayed by $mos4_{KO}$ and $Atcdc5_{KO}$.

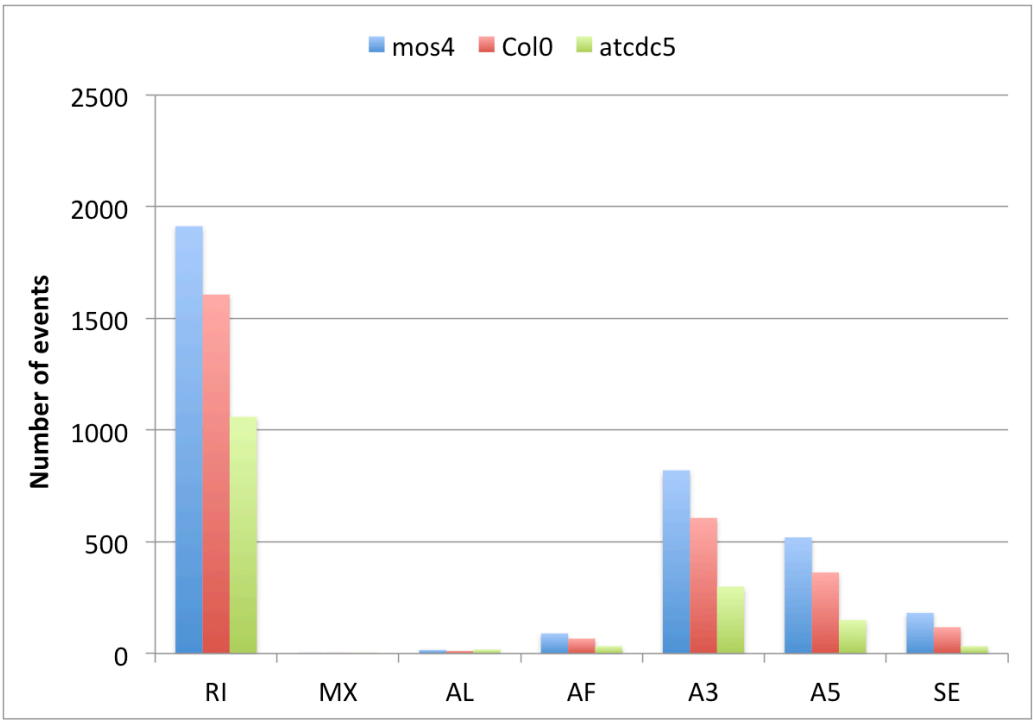


Figure 90. Loss-of-function MAC mutants affect the number of *B. cinerea*-mediated DAS events detected. An increase in the amount of *B. cinerea*-mediated DAS is seen in *mos4*$_{KO}$ mutants compared to wild-type (Col0), whilst fewer DAS events are detected in *Atcdc5*$_{KO}$. However, the trend of RI being the most abundant event type, followed by A3SS (A3), then A5SS (A5) with only a few SE events being detected is seen in all lines.

5.2.7.5 The majority of B. cinerea-mediated DASed genes are affected by loss-of-function of Atcdc5$_{KO}$ or mos4$_{KO}$ transgenic plants

There is very little overlap between the *B. cinerea*-mediated DAS events in either the *Atcdc5*$_{KO}$ or *mos4*$_{KO}$ loss-of-function transgenic plants, with those detected in the wild-type (Col0) lines (Figure 91). Only 3% of DAS events are found in all three lines, which is considerably less than the overlap found when comparing DEG (23.69% of DEG were found in all three lines). This indicates that loss of *Atcdc5*$_{KO}$ and *mos4*$_{KO}$ has a major impact on the DAS seen in response to *B. cinerea*, potentially suggesting that these genes are key nodes in a DAS defence network, however further work would be required to demonstrate this.

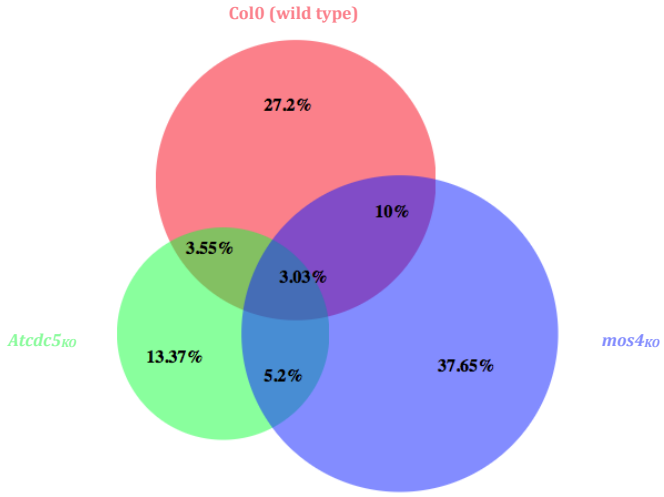


Figure 91. Venn diagram showing the number of genes undergoing DAS events in response to *B. cinerea* 24 hpi in tested lines.

Here we focus in on only those genes that undergo DAS in wild-type plants. Of these 6.9% of Col0 genes underwent *B. cinerea*-mediated DAS in both MAC loss-of-function lines, this corresponds to the 3% mentioned above. However, the majority (62.1%) were only detected to undergo *B. cinerea*-mediated DAS in the wild-type. This indicates that the MAC could be involved in regulating the *B. cinerea*-mediated DAS. To further investigate this, the DAS genes detecting in our dataset was compared with those detected by Cooke (2012). The DAS events that occur in all three lines (Col0, *Atcdc5*$_{KO}$ and *mos4*$_{KO}$) contain a higher percentage of genes that were also detected to undergo *B. cinerea*-mediated DAS in the Cooke dataset.

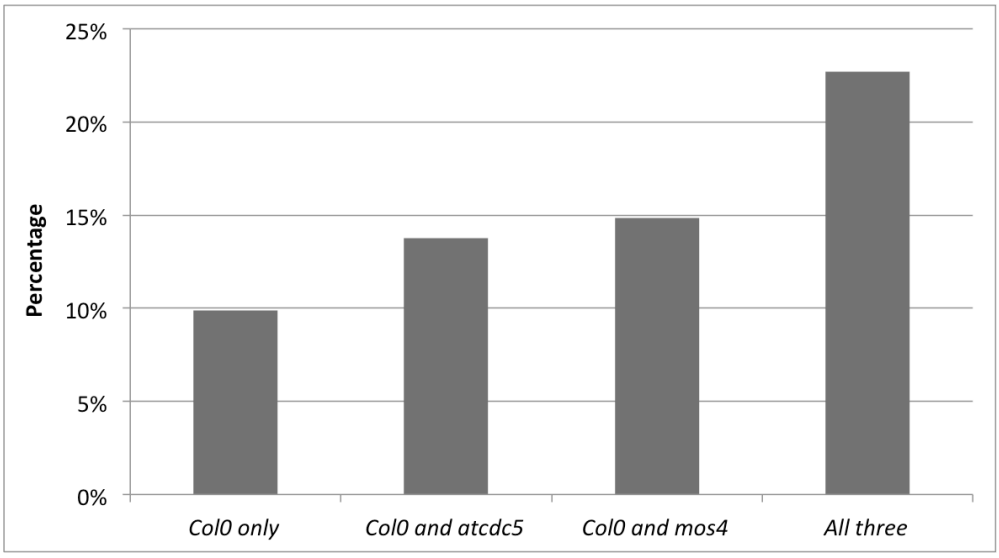


Figure 92. Genes detected to undergo *B. cinerea* mediated DAS in wild-type but not MAC loss-of-function mutants have a smaller overlap with the Cooke 2013 *B. cinerea*-mediated DAS dataset. Dividing the Col0 *B. cinerea*-mediated DAS dataset into those that are uniquely detected in Col0, those that are detected in Col0 and $Atcdc5_{KO}$, those that are detected in Col0 and $mos4_{KO}$, and those that are detected by all three lines, shows that the more lines the DAS event was detected in the larger the amount of overlap there is with the Cooke 2013 dataset.

This indicates that DAS of these genes may be easier to detect, i.e. the magnitude of change of the AS event between infected and non-infected samples may be greater. To further investigate this the magnitude of change, measured by the absolute ΔPSI (|ΔPSI|), of each intersection of the Col0 dataset with the MAC mutants is compared. Here we see that for the majority of the subsections there is no significant difference in the |ΔPSI|, with the exception of $mos4_{KO}$, where the |ΔPSI| is less (ANOVA, $p>0.05$). This indicates that the magnitude of change of the DAS event in the wild-type is not affecting whether it is detected to undergo *B. cinerea*-mediated DAS in the mutant lines.

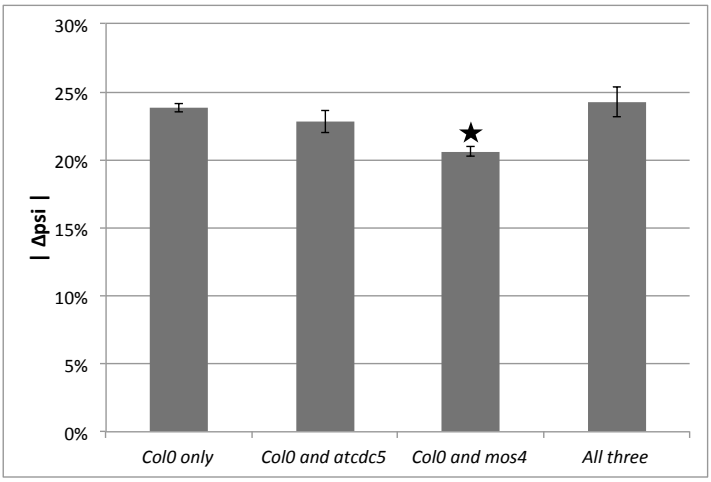


Figure 93. Comparison of the |ΔPSI| of *B. cinerea*-mediated DAS events detected in only Col0 lines, both Col0 and $atcdc5_{KO}$, Col0 and $mos4_{KO}$ and in all three lines. Values represent the average of three replicates±SEM. Asterick indicate statistically different from wild-type (one way anova p<0.05).

5.2.7.6 Genes uniquely detected to undergo B. cinerea mediated differential alternative splicing in the wild-type contain defence-related genes

Mining the list of genes that are only DASed in the wild-type plants (i.e. DAS is abolished in both loss-of-function MAC mutants) shows that the *B. cinerea*-mediated DAS of some defence-related genes are abolished in the MAC mutants. This includes genes associated with the detection and mediation of pathogen responses, such as BAK1, BRI1 and the MAP kinase MKS1, those involved JA signalling, for example JAR1 and NINJA, through to defence related transcription factors, with members of the WRKY, NAC, MYB and EIN families all showing *B. cinerea*-mediated DAS only in the wild-type lines. This indicates that components of the MAC may be regulating the *B. cinerea*-mediated DAS (either directly or indirectly) of defence-related genes. However as mentioned above, independent validation of altered DAS patterns would be required to confirm this.

To further investigate the impact of the MAC on *B. cinerea* defence related splicing we take a closer look at how some of the gene families investigated in "Chapter 4: Genome-wide analysis of Botrytis cinerea-mediated differential alternative splicing in Arabidopsis leaves" are affected.

5.2.7.7 B. cinerea-mediated differential alternative splicing of defence genes associated with the jasmonic acid pathway are affected in Atcdc5$_{KO}$ or mos4$_{KO}$ transgenic plants

In chapter 4 DAS of JAZ genes was analysed. We showed that *B. cinerea*-mediated DAS of *JAZ10* occurs. In response to infection in Col0 wild-type plants the percentage of *JAZ10.4* decreases, whilst the proportion of *JAZ10_ID3* increases, potentially facilitating the fine-tuning of the jasmonate-signalling pathway. This is seen in the *Atcdc5$_{KO}$* transgenic plants. However, it does not occur in the *mos4$_{KO}$* transgenic plants where only a negligible change of the proportion of isoforms is seen (Figure 94).

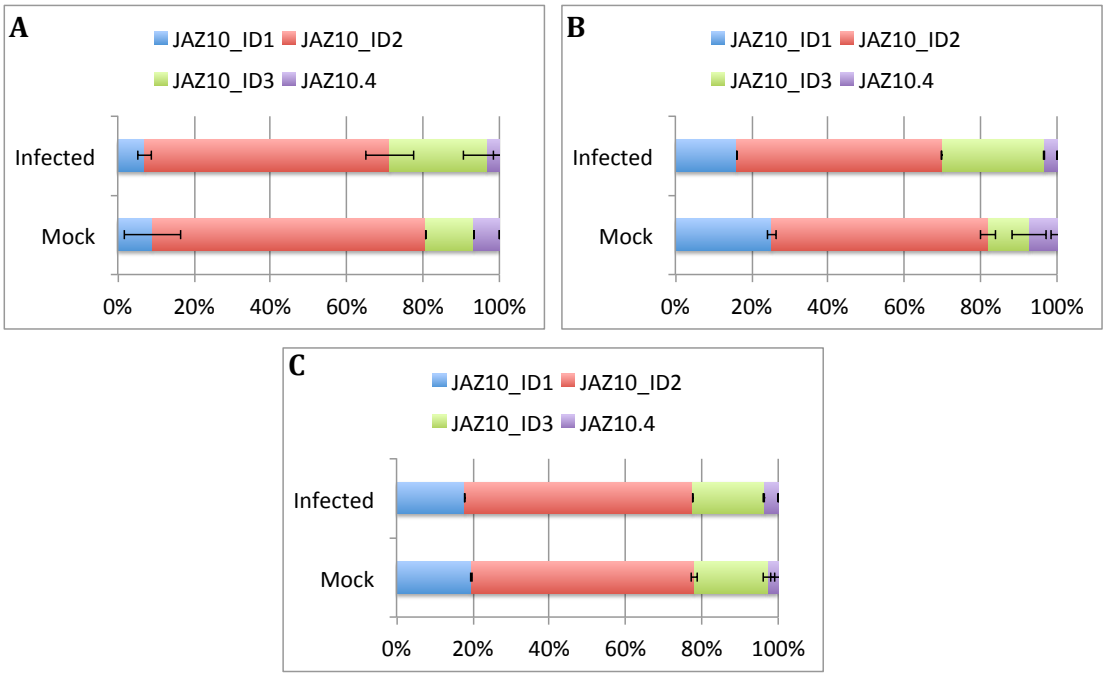


Figure 94. Differential alternative splicing of *JAZ10* is abolished in the loss-of-function mutant *mos4$_{KO}$*. In response to infection the wild-type (A) and the *Atcdc5$_{KO}$* (B) transgenic plants display DAS of *JAZ10*, whereby the percentage of the transcript *JAZ10.4* decreases whilst *JAZ10_ID3* increases. This is not seen in the *mos4$_{KO}$* mutant, where only a negligible change in isoform proportions is observed (C). Error bars represent the SEM of three biological replicates.

This suggests that in the Col0 and *Atcdc5$_{KO}$* lines *B. cinerea*- mediated DAS can help fine-tune the jasmonate-signalling pathway. This ability may be impaired in the *mos4$_{KO}$* lines due to the abolishment of the *B. cinerea* DAS of JAZ10, which could

result in the defence response to *B. cinerea* being affected; this would however require further work to clarify if this is the case.

To further investigate if *B. cinerea*-mediated DASed genes that are associated with the jasmonic acid (JA) defence response are affected by the loss-of-function of $Atcdc5_{KO}$ or $mos4_{KO}$, the DAS of *JAR1* and *NINJA* were examined. JAR1 catalyses the formation of the biologically active jasmonate (JA-Ile), while NINJA acts as a negative regulator of jasmonate responses. We have identified that both of these genes undergo DAS but not DGE in wild-type plants in response to *B. cinerea* infection. In both the $Atcdc5_{KO}$ and $mos4_{KO}$ loss-of-function transgenic lines this DAS is not statistically significant (Students t test p>0.05) (Table 26).

Table 26. Amount of change of percentage spliced in of events associated with Jasmonate signalling pathway that were differentially alternatively spliced in wild-type (Col0) plants. Column one gives gene name, with the event and co-ordinated being DAS given in columns two and three. Columns four and five give the p-value and change of percentage spliced in (ΔPSI) for Col0, with columns six and seven, and eight and nine, giving the values for $Atcdc5_{KO}$ and $mos4_{KO}$ respectively. Statistical significance was calculated using the SUPPA software, by comparing the observed ΔPSI between conditions with the distribution of the ΔPSI between replicates as a function of the gene expression. P values are given to two decimal places, with the ΔPSI is given to one decimal place.

			Col0		$Atcdc5_{KO}$		$mos4_{KO}$	
Gene	Event	Co-ordinates	p-value	ΔPSI	p-value	ΔPSI	p-value	ΔPSI
JAR1	A3SS	Chr2:19034809-19034956: 19034809-19035273	0.04	12.6%	0.16	3.4%	0.15	-1.4%
NINJA	A3SS	Chr4:14266144-14266536: 14266105-14266536	0.09	-5.2%	0.19	-2.5%	0.12	-2.6%
	RI	Chr4:14264703-14264789	0.04	-19.4%	0.12	-6.3%	0.11	8.3%
	SE	Chr4:14265757-14266023: 14266105-14266536	0.05	17.7%	0.16	6.7%	0.09	9.7%
	SE	Chr4:14265757-14266023: 14266144-14266536	0.05	13.7%	0.10	13.3%	0.11	3.5%

Taken together these results indicate that at least some components of the MAC may be involved in the regulation of DAS of genes associated with the jasmonate signalling pathway in response to *B. cinerea* infection.

5.2.7.8 Loss-of-function mutants Atcdc5$_{KO}$ and mos4$_{KO}$ affect the differential splicing of members of the receptor like kinase family

To help to determine if other types of defence responses are likely to be influenced by DAS regulated by components of the MAC, the RLK family of genes was investigated. This is because its members play a distinctly different role in the defence response (primarily detection of pathogens), and we have previously shown in Chapter 4 that a number of these genes undergo *B. cinerea*-mediated DAS.
Here we will investigate three RLK genes that encode proteins associated with the perception of chitin, a classical fungal PAMP, that we have identified to undergo *B. cinerea*-mediated DAS in wild-type Col0 plants. These three RLKs are; CERK1 a well-known pattern recognition receptor (PRR) for the chitin (Wan *et al*., 2008), PBL27, which is immediately downstream of CERK1 and has been implicated to play a role in modulating chitin-induced immunity (Shinya *et al*., 2014), and BAK1, a proposed modulator of the plant defence responses (Roux *et al*., 2011).

5.2.7.8.1 Loss-of-function mutants Atcdc5$_{KO}$ and mos4$_{KO}$ affect the differential splicing of CERK1

In wild-type plants *CERK1.1* increases in proportion in response to *B. cinerea* infection whilst *CERK1_ID4*, which we predict to be targeted for NMD decreases. CERK1_ID3, also predicted to be targeted for NMD, accounts for a very small proportion of total transcripts (Figure 95A). In the *Atcdc5$_{KO}$ B. cinerea*-mediated DAS is abolished (Figure 95B). In both *Atcdc5$_{KO}$* and *mos4$_{KO}$*, CERK1_ID3 accounts for a larger proportion of the total transcript amount than in the wild-type, which in the *mos4$_{KO}$* mutant contributes to a reduction in the functional transcript in response to *B. cinerea* infection, and an increase in those transcripts predicted to be target for NMD (*CERK1_ID3* and *CERK1_ID4*) (Figure 95C).

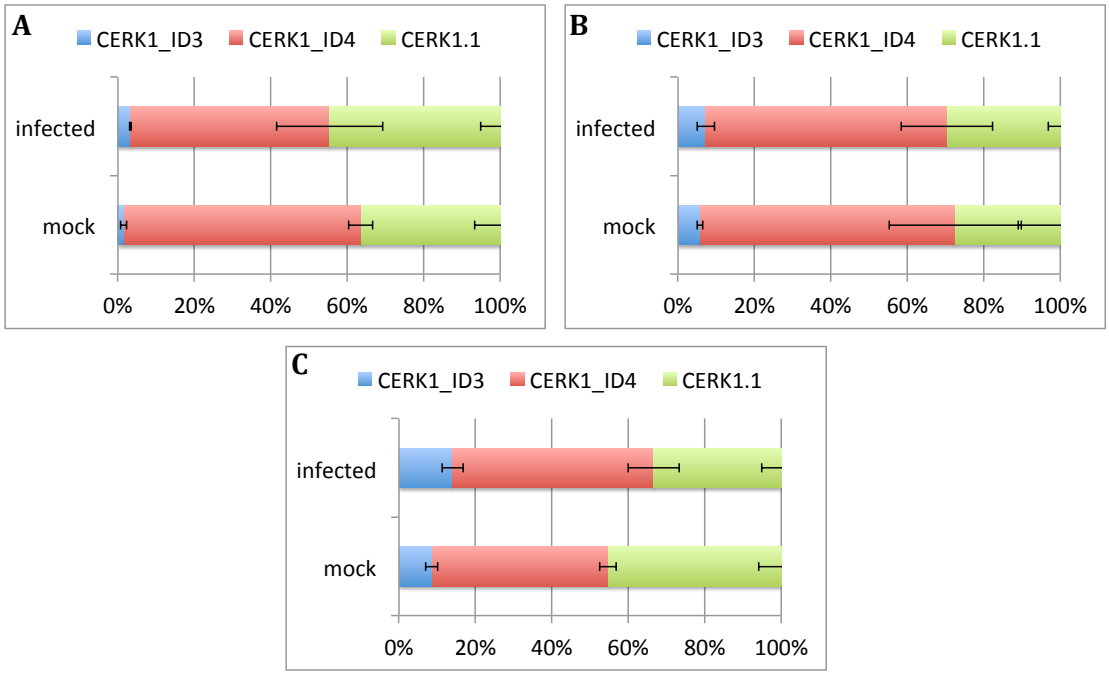


Figure 95. Differential alternative splicing of *CERK1* is affected in the loss-of-function mutants *Atcdc5$_{KO}$* and *mos4$_{KO}$*. The proportion of the three detected CERK1 isoforms from infected (top bar) and mock (bottom bar) are shown for Col0 (A), *Atcdc5$_{KO}$* (B) and *mos4$_{KO}$* (C) lines. Error bars represent the SEM of three biological replicates.

This could potentially indicate a role for *Atcdc5* and *mos4* in DAS of *CERK1* with *Atcdc5* being generally involved in the defence-mediated DAS of *CERK1* and *mos4* potentially being more actively involved in regulating the DAS of isoforms that are coupled to NMD. However, both *Atcdc5* and *mos4* influence the production of alternative isoforms. It is possible that ATCDC5 and MOS4 work together as part of the MAC to bring about promotion or suppression of a splicing event in a co-ordinated manner rather than one being general and the other more specific. Considerable additional work would be required to confirm which of these theories is more plausible.

5.2.7.8.2 Loss-of-function mutants *Atcdc5$_{KO}$* and *mos4$_{KO}$* abolish the differential splicing of PBL27

B. cinerea-mediated DAS of the RLKs immediately downstream of *CERK1, PBL27*, is also altered in *Atcdc5$_{KO}$* and *mos4$_{KO}$* transgenic plants. In wild-type plants DAS but not DGE is observed in response to *B. cinerea*, with the *PBL27.2* isoform being

increased by 23% in response to infection, which we predicted would increase the gene expression in a transcriptional independent manner, due to the RI event in the 3'UTR of this isoform. This DAS is abolished in both the *Atcdc5$_{KO}$* and *mos4$_{KO}$* transgenic plants; DGE is not detected in either of the transgenic lines (the same as for wild-type plants). However, these observations come from an area with low read coverage. To confirm that the *B. cinerea*-mediated DAS does not occur in the MAC KO lines validation of the RI events would be required. We predict that in response to *B. cinerea* infection the amount of the PBL27 protein would increase in wild-type plants due to the DAS RI event in the 3'UTR, by mechanisms discussed in chapter 4 section 4.3.9.1. However due to the *B. cinerea*-mediated DAS not occurring in *Atcdc5$_{KO}$* and *mos4$_{KO}$* transgenic plants, the levels of PBL27 would remain constant. Additional experiments measuring the amount of the PBL27 protein produced in response to *B. cinerea* in these lines would be required to test this hypothesis.

5.2.7.8.3 Loss-of-function mutant *mos4$_{KO}$* affect the differential splicing of *BAK1*

The *B. cinerea*-mediated DAS of *BAK1* is also abolished in *mos4$_{KO}$* transgenic plants. In wild-type plants *BAK1* undergoes DAS and DGE in response to *B. cinerea*, with the functional transcript *BAK1.1* increasing in proportion compared to the *BAK1_ID3* transcript, which we predict to be targeted for NMD due to an A5SS event (Figure 96A). DGE is still observed in response to *B. cinerea* infection in both the *Atcdc5$_{KO}$* and *mos4$_{KO}$* transgenic plants. However the *B. cinerea*-mediated DAS is abolished in the *mos4$_{KO}$* line (Figure 96B). In the *Atcdc5$_{KO}$* line a proportional change of transcript does occur with borderline statistical significance (p=0.05). This is probably due to the large variation of the transcript levels seen in the mock biological replicates. Thus we predict that DAS does indeed occur in this line (Figure 96C). Examining the abundances of each transcript shows that although *B. cinerea*-mediated DAS is not observed in the *mos4$_{KO}$* transgenic plants, it has the highest amount of *BAK1.1* transcript overall (Figure 96D-F). This indicates that loss-of-function of *mos4$_{KO}$* not only abolishes the *B. cinerea*-mediated DAS of *BAK1*, but it also affects the overall expression levels of transcripts. This could potentially indicate that *mos4$_{KO}$* is involved in regulating the expression of *BAK1*, or affects genes involved in the

regulation of BAK1 as well as affecting the DAS. Possibly by *mos4* interacting or affecting NMD mechanisms. Circumstantial evidence that supports this theory is that one of the genes involved in the core NMD machinery (*smg7*) is DE in response to *B. cinerea* in the $Atcdc5_{KO}$ and $mos4_{KO}$ lines but not in the wild-type lines.

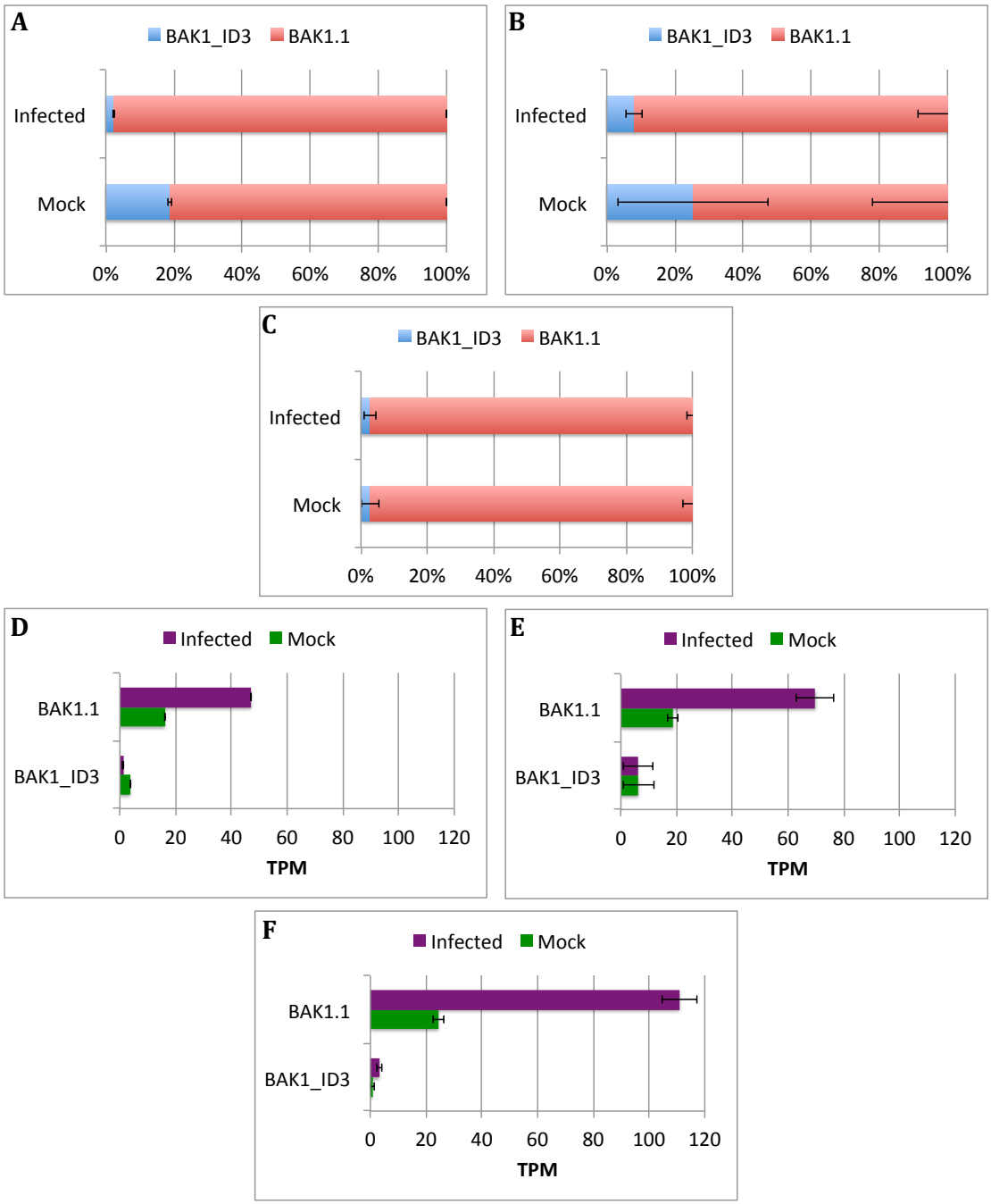


Figure 96. Differential alternative splicing of *BAK1* is affected in the loss-of-function mutants *Atcdc5$_{KO}$* and *mos4$_{KO}$*. The proportion of the two detected *BAK1* isoforms from infected (top bar) and mock (bottom bar) are shown for Col0 (A), *Atcdc5$_{KO}$* (B) and *mos4$_{KO}$* (C) lines. The total abundance of the *BAK1* isoforms from infected (purle) and mock (green) are shown for Col0 (D), *Atcdc5$_{KO}$* (E) and *mos4$_{KO}$* (F) lines. Error bars represent the SEM of three biological replicates.

Taken together these results indicate that both *Atcdc5$_{KO}$* and *mos4$_{KO}$* affect the *B. cinerea* mediated DAS of genes associated with the perception of chitin. This,

combined with the results from the investigation into genes associated with JA signalling indicates that at least some components of the MAC are involved in the regulation of DAS of a range of defence-related genes. Additional evidence for this comes from the list of genes that are only DASed in wild-type plants containing defence-related genes.

As well as directly acting upon defence-related genes, we saw in chapter 4 that splicing factor associated genes, in particular the SR gene family, were highly regulated by DAS in response to *B. cinerea* infection. Comparing the *B. cinerea*-mediated DAS of SR proteins seen in Col0 samples with the $Atcdc5_{KO}$ and $mos4_{KO}$ samples could help us to begin to elucidate how these splicing associated factors interact with each other to bring about and modulate *B. cinerea* mediated DAS. In addition, investigating whether $Atcdc5_{KO}$ or $mos4_{KO}$ mutants affect *B. cinerea*-mediated DAS of other splicing factor associated genes could help to explain the breadth of DAS effects seen in mutants, because affecting DAS of genes encoding SR proteins could lead to cascade effect.

5.2.7.9 MAC loss-of-function mutants can affect the B. cinerea-mediated differential alternative splicing of genes that encode SR proteins

In chapter 4, we determined that the majority (14 out of 18) of the genes that encode Arabidopsis SR proteins undergo DAS in response to *B. cinerea* 24 hpi. However, at the event level (the analysis we are utilising here) only five are DASed at the 5% FDR level. Interestingly the $mos4_{KO}$ mutants appear to have increased splicing of SR proteins with the majority (11 out of 14) of SR protein encoding genes detected to undergo *B. cinerea*-mediated DAS with a FDR of 5%, while the $Atcdc5_{KO}$ detect fewer gene as DAS (Table 27).

Table 27. *B. cinerea*-mediated DAS of SRprotein encoding genes in the *Atcdc5$_{KO}$* and *mos4$_{KO}$* transgenic plants. Column one and two give the gene identifier and gene name respectively. Column threeand four, five and six, and, seven and eight give the p-value and ΔPSI associated the most significant change of percentage spliced in of an event of this gene in Col0, *Atcdc5$_{KO}$* and *mos4$_{KO}$* transgenic plants respectively. Green, highlight values where p<0.05.

		Col0		Atcdc5$_{KO}$		mos4$_{KO}$	
Gene Identity	Gene Name	p-value	ΔPSI	p-value	ΔPSI	p-value	ΔPSI
AT1G02840	SR34	0.04	-12.3%	0.04	-28%	0.03	-17.6%
AT1G09140	SR30	0.04	-13.2%	0.03	-27.5%	0.00	-42.1%
AT1G55310	SCL33	0.06	6.7%	0.09	-9.1%	0.01	-20.5%
AT2G24590	RSZ22a	0.09	-3.4%	0.25	0.3%	0.03	12.5%
AT2G37340	RSZ33	0.02	23.4%	0.06	15.1%	0.04	16%
AT2G46610	RS31A	0.06	12.7%	0.04	-31.2%	0.05	24.7%
AT3G13570	SCL30a	0.07	8%	0.14	-5.8%	0.01	-33.9%
AT3G49430	SR34a	0.12	-4.5%	0.03	-3.6%	0.06	-7%
AT3G53500	RSZ32	0.11	2.3%	0.11	5.4%	0.06	-6.4%
AT3G61860	RS31	0.05	7.9%	0.10	-7.8%	0.04	-9.4%
AT4G02430	atSR34b	0.04	12.2%	0.06	15%	0.07	12.2%
AT4G25500	RS40	0.07	9%	0.10	-11.2%	0.02	-19.3%
AT4G31580	RSZ22	0.06	-7.7%	0.10	-8.8%	0.03	-11.6%
AT5G52040	RS41	0.04	11.2%	0.08	20.3%	0.04	-11.2%

It is possible that not detecting SR encoding genes undergoing *B. cinerea*-mediated DAS in the *Atcdc5$_{KO}$* could be due to an issue with sensitivity or power of the analysis rather than a real biological phenomenon. Comparing the DAS of the 12 SR encoding genes that display *B. cinerea*-mediated ΔPSI of events with a p-value <0.1 in Col0 plants indicates that there is not a straight forward relationship between loss-of-function of AtCDC5 or MOS4 and how the *B. cinerea*-mediated DAS of a SR protein encoding gene is affected.

The gene encoding SR34 undergoes three different *B. cinerea*-mediated DAS events in the wild-type. An A3’ event (628469-628935:628469-629288) that is significant at the 5% level (p-value = 0.04), this event is also detected to be DAS in both of the MAC loss-of-function mutants. At the 10% level two other events, an A3’ (626828-626913:626828-626917) and SE event (628469-628777:628856-628935) are detected, with p values of 0.06 and 0.09 respectively. These two DAS events are

altered in the MAC loss-of-function mutants. In infected samples the A3' event (626828-626913:626828-626917) is up-regulated in wild-type plants, whereas in both MAC loss-of-function mutants it is down-regulated. This is also the case for the SE event in the *mos4*$_{KO}$ mutant. However the *B. cinerea*-mediated DAS is abolished in the *Atcdc5*$_{KO}$ (Figure 97). Further work would be required to determine if these DAS events are occurring in the wild-type and to confirm if they are altered in the loss-of-function mutants.

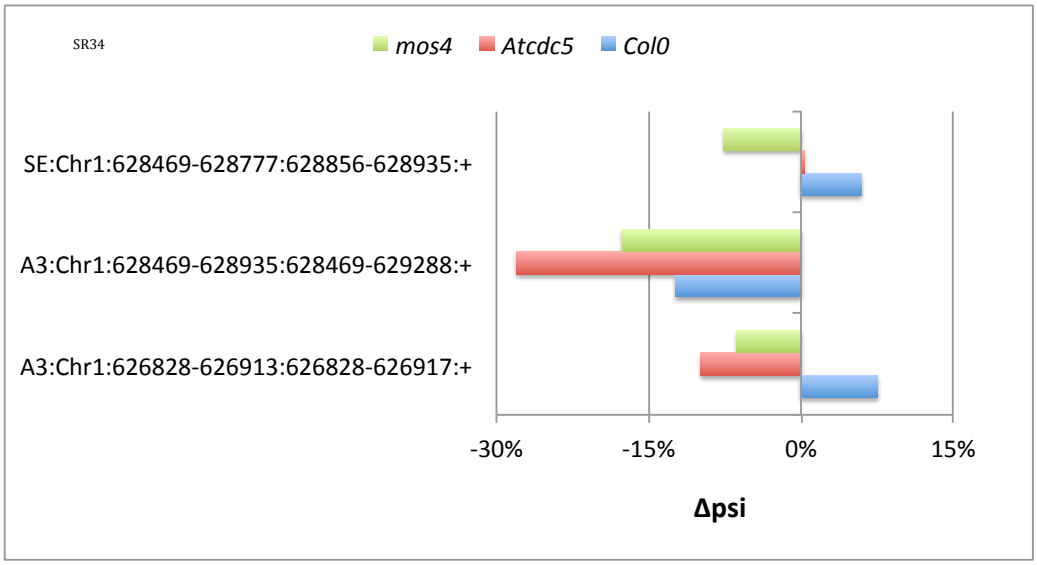


Figure 97. Three potential *B. cinerea*-mediated differential alternatively spliced events of *SR34.* Events are detailed on the left-hand side. Bars represent the ΔPSI in *mos4*$_{KO}$ (green), *atcdc5*$_{KO}$ (red) and Col0 (blue) lines.

Further evidence that in *Atcdc5*$_{KO}$ lines *B. cinerea*-mediated DAS is severely disrupted comes from investigating *RS41* and *RSZ22a*. In *RS41*, the amount of the A5SS event increases in wild-type plants, whereas it decreases in the *mos4*$_{KO}$ and no change is detected in the *Atcdc5*$_{KO}$, i.e. *B. cinerea*-mediated DAS is abolished (p= 0.04, 0.12, and 0.04 respectively) (Figure 98, top panel). Whereas the DASed RI event of *RSZ22a* is decreased in wild-type plants but increased in *mos4*$_{KO}$ lines, the DAS is again seen to be abolished in the *Atcdc5*$_{KO}$ lines (p=0.08, 0.25 and 0.03 for the Col0, *Atcdc5*$_{KO}$ and *mos4*$_{KO}$ lines respectively) (Figure 98, bottom panel).

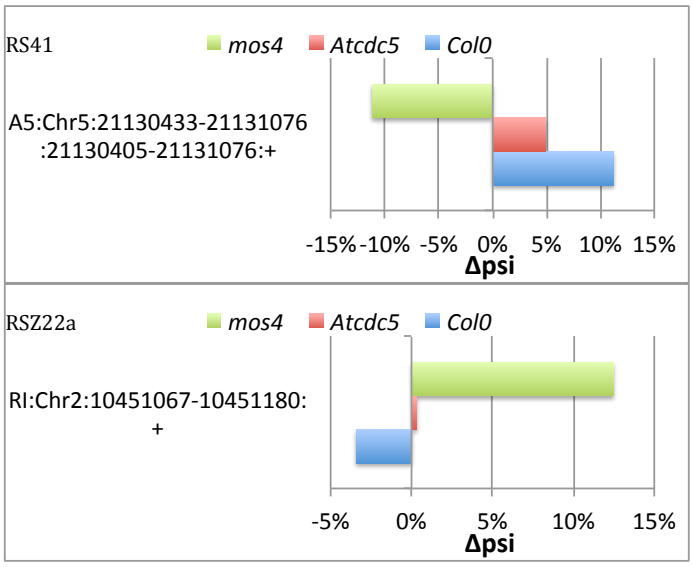


Figure 98. *B. cinerea*-mediated DAS of *RS41* and *RSZ22a* is abolished in *Atcdc5*$_{KO}$ lines. Events are detailed on the left-hand side. Bars represent the ΔPSI in *mos4*$_{KO}$ (green), *atcdc5*$_{KO}$ (red) and Col0 (blue) lines.

This is the type of response could potentially indicate that MOS4 represses defence related splicing (hence when it is removed splicing is enhanced) whilst AtCDC5 activates splicing. However, investigating other SR encoding genes that undergo *B. cinerea*-mediated DAS indicates that the picture is more complex. In two genes that encode members of the SCL family, SCL33 and SCL30a, the *B. cinerea*-mediated DAS is abolished in the *Atcdc5*$_{KO}$. However, in this case the DAS is unaltered in the *mos4*$_{KO}$ (Figure 99). This indicates that there is not a straightforward regulatory relationship where AtCDC5 activates and MOS4 represses defence mediated splicing. Additional evidence for this comes from examining the effects of the MAC loss-of-function mutants on the *B. cinerea*-mediated DAS patterns of *SR30*. For the A5SS event *B. cinerea*-mediated DAS is abolished in the *mos4*$_{KO}$ mutant while in the *Atcdc5*$_{KO}$ it is enhanced, however in the A3SS event both loss-of-function mutants have increased ΔPSI compared to that seen in wild-type lines (Figure 100).

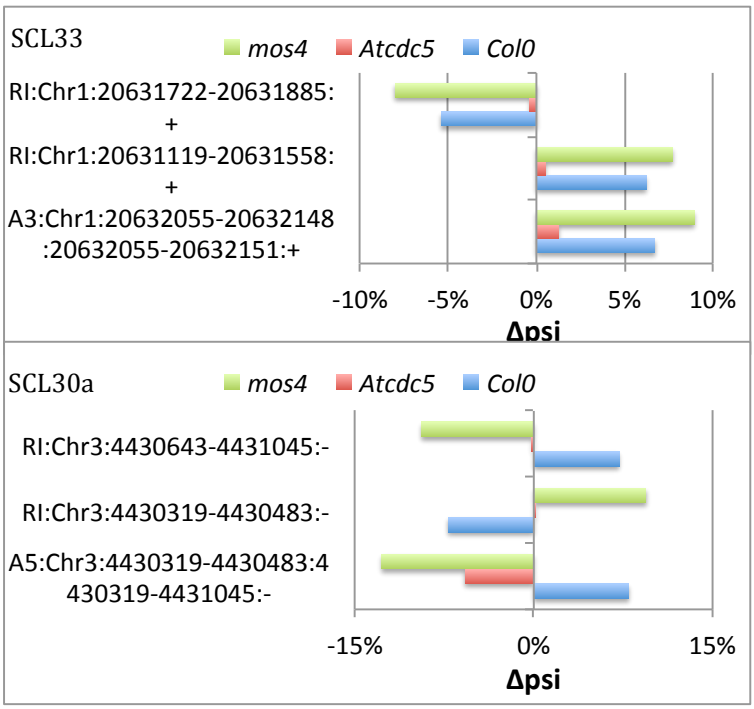


Figure 99. *B. cinerea*-mediated DAS of *SCL33* and *SCL30a* is abolished in *Atcdc5*$_{KO}$ lines but unaltered in *mos4*$_{KO}$. Top panel. *SCL33,* bottom panel *SCL30a*. Events are detailed on the left-hand side. Bars represent the ΔPSI in *mos4*$_{KO}$ (green), *atcdc5*$_{KO}$ (red) and Col0 (blue) lines.

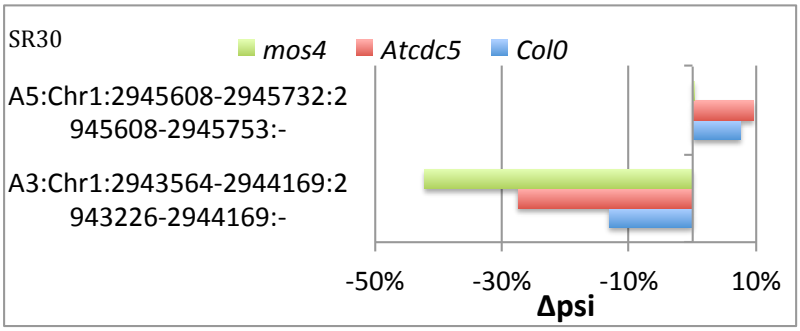


Figure 100. *B. cinerea*-mediated DAS of SR30 is abolished in the loss-of-function *mos4*$_{KO}$. Events are detailed on the left-hand side. Bars represent the ΔPSI in *mos4*$_{KO}$ (green), *atcdc5*$_{KO}$ (red) and Col0 (blue) lines.

Taken together these results show that the *B. cinerea*-mediated DAS of some genes encoding SR proteins is affected in the loss-of-function mutants *Atcdc5*$_{KO}$ and *mos4*$_{KO}$; this could potentially suggest that they are involved in the regulating the splicing of these genes. However, the effect on the DAS in the loss-of-function mutants is not the same for all SR encoding genes. This indicates that there is a more complex regulatory network than AtCDC5 or MOS4 generally suppressing or

enhancing *B. cinerea*-mediated DAS. To investigate this further other splicing factor associated genes were investigated.

Comparing the proportions of the families of splicing factor associated genes, shows that the largest family affected in all lines is the hnRNP protein gene family. However this is reflected in the fact that they represent a higher proportion of the total number of splicing factor associated genes investigated. It is also highlights the fact that no genes that encode components of the MAC undergo *B. cinerea*-mediated DAS in all three lines (Figure 101). This could potentially indicate that the complex itself is involved in splicing components of complex.

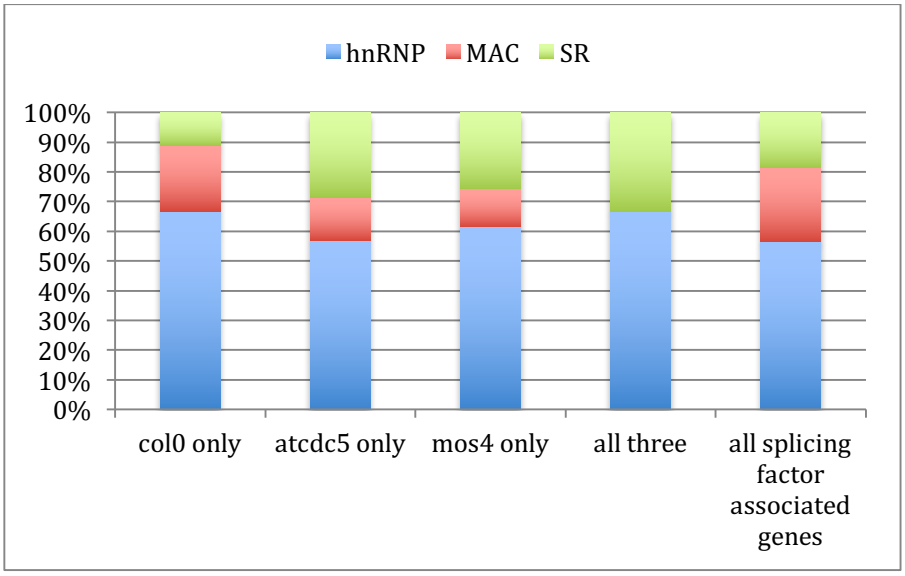


Figure 101. Comparison of the proportion of DASed genes from three main splicing factor associated gene families. Values represent the percentage of genes encoding hnRNP, MAC and SR proteins that are DASed uniquely in one line (bars 1-3) and in all three lines (bar four). Bar five represents the proportion present in the whole Arabidopsis genome.

To further investigate this the DAS of genes that encode two functionally redundant components of the MAC, *MAC3A* and *MAC3B*, that Monaghan *et al* 2009 showed to be involved in the regulation of plant innate immunity are examined. *MAC3B* was uniquely DAS in the wild-type lines at the 5% FDR level, the MAC loss-of-function mutants are not detected to undergo statistically significant DAS. The A5SS DAS event of *MAC3A* is not significant in wild-type lines but in the MAC loss-of-function

mutant *mos4*$_{KO}$ this event undergoes a statistically significant (p=0.02) decrease in ΔPSI by 21.6% (

Table 28). This indicates that the *B. cinerea*-mediated DAS of these genes is at least in part regulated by the MAC, highlighting that this spliceosomal associated complex can regulate its components via DAS. This could potentially be acting as a feedback mechanism, however further work would be required to test this hypothesis.

Table 28. *B. cinerea*-mediated DAS events of *MAC3A* and *MAC3B*. Column one gives the gene name with the DAS event depicted in column two. Columns 3 and 4, 5 and 6, and 7 and 8 give the p-value and ΔPSI for Col0, *Atcdc5*$_{KO}$ and *mos4*$_{KO}$ respectively. Green cells highlight values where p<0.05.

		Col0		*Atcdc5*$_{KO}$		*mos4*$_{KO}$	
Gene	Event	p-value	ΔPSI	p-value	ΔPSI	p-value	ΔPSI
MAC3A	AT1G04510;A5:Chr1:1227957-1228046:1227948-1228046	0.07	9.5%	0.15	-5.2%	0.02	-21.6%
MAC3A	AT1G04510;A3:Chr1:1229977-1230065:1229977-1230070	0.22	0.0%	0.25	0.0%	0.16	-1.3%
MAC3B	AT2G33340;A5:Chr2:14130892-14130988:14130892-14130992	0.02	21.2%	0.07	-12.4%	0.10	-3.2%
MAC3B	AT2G33340;RI:Chr2:14126492-14126577	0.22	-0.1%	0.23	0.4%	0.22	0.4%

These DAS events correspond to an increase of the *MAC3B.1* isoform accompanied by a decrease in *MAC3B_ID20* isoforms in response to *B. cinerea* infection in wild-type plants (Figure 102).

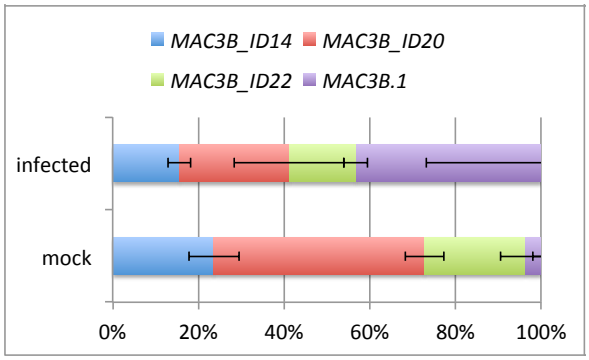


Figure 102. *B. cinerea*-mediated DAS of *mac3b* increase the proportion of the functional transcript.

The A5SS in *MAC3B_ID20* results in a PTC due to a frame shift, which we predict will target the transcripts for NMD. Thus in response to infection, DAS results in an increase in the functional transcripts; this potentially enables the plant to respond faster to the pathogen attack by circumventing the time required for the initiation of transcription and mRNA accumulation.

This work on the splicing factor associated genes indicates that there is a complex interwoven network of splicing factors as shown by SR proteins and components of the MAC having altered DAS in the loss-of-function mutants compared to the wild-type lines.

5.3 Discussion

5.3.1 Loss-of-function mutants of six components of the MAC display *B. cinerea* defence phenotypes

We have identified six loss-of-function MAC mutants that have increased susceptibility to *B. cinerea*, four of which also displayed altered morphology. As first observed by Palma *et al* (2007) and Monaghan *et al* (2012), $mos4_{KO}$, $Atcdc5_{KO}$ and $prl1_{KO}$ were smaller than Col0 wild-type plants and had smaller leaves, and $mac5a_{KO}$ had slightly elongated leaves that often appeared twisted. This correlated with our findings that these four loss-of-function mutants seedlings had reduced root length, with $mos4_{KO}$, $Atcdc5_{KO}$ and $prl1_{KO}$ also being lighter than wild-type plants. Because those lines with altered morphology are either backcrossed lines, or in the case of $mac5a_{KO,}$ two independent T-DNA lines displaying the same altered morphology this strongly implies that the defects are due to the gene that we are targeting. This indicates that these genes are highly likely to also be involved in processes other than plant defence.

Of the six MAC loss-of-function mutants that displayed increased susceptibility to *B. cinerea*, $mac5a_{KO}$ and $At3g20820_{KO}$ did not display increased susceptibility to *Pst*; Monaghan *et al* (2010) also states that $mac5a_{KO}$ is not more susceptible to *Pst* potentially due to partial redundancy with *MAC5B*. This suggests that these genes could potentially play a role specifically in fungal or necrotrophic pathogen defence. Investigating $mac5a_{KO}$ and $At3g20820_{KO}$ loss-of-function mutants susceptibility to additional pathogens could help to determine if it is specific to fungal or necrotrophic pathogens, helping to identify potential splicing elements involved specifically in response to subsets of pathogens.

5.3.2 Components of the MAC are involved in *B. cinerea*-mediated differential alternative splicing

Although the MAC has previously been shown to be involved in the plant defence response, to our knowledge this is the first time that the effect of components of the MAC have been investigated regarding their effect on the splicing patterns of genes at the genome-wide

scale; previously AtCDC5 and MOS4 have been shown to contribute to the proper splicing of two R genes, *SNC1* and *RPS4* (F. Xu *et al*., 2012).

A screen of the *B. cinerea* susceptible loss-of-function mutants showed that *AT4G39270* underwent *B. cinerea*-mediated DAS in all apart from $Atcdc5_{KO}$, potentially indicating that ATCDC5 plays a role in *B. cinerea*-mediated splicing. Further evidence of this came from the genome-wide study that showed loss of *MOS4* or *ATCDC5* affects a plethora of transcripts across a range of defence and splicing related families; with 15.9% of RLK and 30% of splicing factor associated genes having altered *B. cinerea*-mediated DAS in one or both loss-of-function transgenic lines. We have shown that the $Atcdc5_{KO}$ and $mos4_{KO}$ alter the DAS patterns of a variety of other splicing factor genes including genes encoding SR proteins and other components of the MAC. It is possible that the wide breadth of altered DAS observed is due to a cascade effect, i.e. altering the splicing pattern of a splicing factor associated gene would alter its ability to splice other genes.

5.3.3 $Atcdc5_{KO}$ and $mos4_{KO}$ affect a wide range of splicing factor-associated genes

Genes from each of the splicing-factor-associated families are seen to have altered *B. cinerea*-mediated DAS in each of the lines investigated; Col0, $Atcdc5_{KO}$ and $mos4_{KO}$. However, no genes that encode components of the MAC underwent *B. cinerea*-mediated DAS in the same manner in all three lines, strongly indicating that ATCDC5 and MOS4 contribute to regulating the differential splicing of other members of the MAC much in the same way as SR protein family members have been shown to regulate multiple other SR protein genes.

Genes encoding the SR proteins also have altered *B. cinerea*-mediated DAS in the $Atcdc5_{KO}$ and $mos4_{KO}$ transgenic plants compared to wild-type plants, indicating that they too are in part regulated by components of the MAC. In mice, Pandit *et al* (2013) determined that SR proteins are in direct competition with other nearby SR proteins and splicing regulators for each specific binding event, and that when they removed a specific SR protein the complex response they observed was due to not just to the loss of function of that SR protein, but due to coordinated modification of binding, and thus splicing activity, of nearby splicing

regulators. Our work on the MAC indicates that a similar scenario could be occurring here, where the massive change in *B. cinerea* DAS seen in the loss-of-function *Atcdc5$_{KO}$* or *mos4$_{KO}$* is not due solely to the loss of function of this protein, but is associated with a multitude of interconnected splicing factors working in conjunction to bring about DAS. Thus perturbation of this network results in a massive change of DAS patterns in a plethora of genes.

This could explain the breadth of gene effect but the question remains, how do components of the MAC alter splicing patterns? In HeLa cells Song *et al* (2010) have shown that the PRP19 (human equivalent of the MAC) affects the stability of the U4/U6.U5 component of the spliceosome by ubiquitin-dependant regulation, potentially facilitating interactions between spliceosome components and rearrangement of the spliceosome. It is conceivable that a similar situation occurs in plants with the MAC and that potentially this ubiquitin-dependant regulation could alter splice site choice. Although very hypothetical, this could also explain why we see overlapping but different effects of *Atcdc5$_{KO}$* and *mos4$_{KO}$* on DAS patterns.

5.3.4 Components of the MAC affect *B. cinerea*-mediated DAS of defence-related genes at multiple levels of the defence pathway

Evidence that components of the MAC could potentially be involved in DAS that plays a role at multiple levels of the defence pathways was gained by mining those genes that only undergo *B. cinerea*-mediated DAS in wild-type lines. These genes are likely to be regulated at least in part by AtCDC5 and MOS4, because without them DAS is abolished. This dataset contains genes that represent multiple stages in the defence pathway from detection of pathogen, to transduction of signalling to initiation of defence-related genes.

5.3.5 Three receptor like kinases involved in the perception of *B. cinerea* have altered *B. cinerea*-mediated DAS in Atcdc5$_{KO}$ and mos4$_{KO}$ transgenic plants

The gene encoding the cell surface receptor of chitin, *CERK1*, and a component immediately downstream that regulate chitin induced immunity is DASed in wild-type plants. In chapter 4

we proposed that this *B. cinerea*-mediated DAS could affect expression levels of CERK1 and PBL27. Both the *Atcdc5*$_{KO}$ and *mos4*$_{KO}$ affect the DAS of these genes, with the DAS of PBL27 being abolished in both transgenic lines. We hypothesis that because PBL27 does not undergo DGE this would result in mock and infected samples containing the same amount of proteins, whereas the Col0 wild-type samples would have elevated levels due to the DAS event. However this would require experimental validation. Interestingly the DAS of the RLK *BAK1*, which has been described as a modulator of plant defence responses, is only affected by *mos4*$_{KO}$, once again indicating that *ATCDC5* and *MOS4* have divergent but overlapping regulatory roles in regards to plant defence.

5.3.6 Loss-of-function MAC transgenic Arabidopsis plants lose the ability to differentially alternatively splice jasmonic acid signalling associate genes JAZ10, JAR1 and NINJA

In wild-type Col0 plants three genes known to be associated with the JA signalling pathway, *JAZ10*, *JAR1* and *NINJA* are DASed in response to *B. cinerea*, suggesting that DAS is involved in regulating JA signalling in pathogen defence responses. The *B. cinerea*-mediated DAS of *JAZ10* is abolished in loss-of-function *Atcdc5*$_{KO}$ mutants, but is present in *mos4*$_{KO}$ mutants. This potentially indicates that *Atcdc5*$_{KO}$ is involved in the regulation of the negative feedback of JAZ10, because in chapter 4 we hypothesised that the *B. cinerea*-mediated DAS would affect the negative feedback mechanisms of JAZ10. Two other genes associated with JA signalling, *JAR1* and *NINJA* show *B. cinerea*-mediated DAS in wild-type plants, with this DAS being abolished in both the *Atcdc5*$_{KO}$ and *mos4*$_{KO}$ transgenic lines.

JAR1 is involved in the formation of JA-Ile, and in response to *B. cinerea* infection in wild-type plants there is a significant reduction in the isoform that contains a PTC due to an A3SS; we hypothesis that this PTC would result in this isoform being targeted for NMD. Thus in response to infection, the *B. cinerea*-mediated DAS could induce a rapid response resulting in increased protein levels because it could circumvent the time required for transcriptional activation and mRNA accumulation. This could in turn result in increased formation of JA-Ile, cumulating in the activation of JA responsive genes. This DAS is abolished in both the *Atcdc5*$_{KO}$ and *mos4*$_{KO}$ transgenic lines. Thus, we hypothesise that one of the results of

knocking out these genes could be a slower accumulation of activated JA in response to *B. cinerea* infection.

B. cinerea-mediated DAS of *NINJA* occurs in wild-type lines. *NINJA_ID6* has a RI that contains a PTC that we predict will result in a truncated protein that if translated can not interact with TOPLESS (TPL). In response to *B. cinerea* infection there is a reduction in the RI transcript, which we predict would result in more functional protein being produced. Why this is the case and what is it biological significance is unknown. However it has been shown in R genes that both full length and truncated versions of the proteins are required to confer resistance to pathogens, with Zhang and Gassman (2003) proposing that this may be due to the truncated protein alleviating self-inhibition. A similar mechanism could potentially be occurring here. Further work to determine if AS does produce truncated NINJA proteins and what their role in *B. cinerea* infection is could help to further expand our knowledge of how transcription is regulated in response pathogen attack. This DAS was only detected in wild-type lines potentially, indicating that AtCDC5 and MOS4 could be involved in modulating this response.

Taken together this work indicates that there are putative roles for components of the MAC being involved in regulating the JA signalling pathway at multiple stages in the pathway.

5.3.7 A putative role for components of the MAC regulating transcription factor is identified

Multiple transcription factors with members from the WRKY, NAC, MYB and EIN families undergo *B. cinerea*-mediated DAS only in the wild-type lines. This indicates that components of the MAC may be involved in their regulation, because if either AtCDC5 or MOS4 is not present their DAS is abolished. In rice Liu *et al* (2016) have shown that two WRKY transcription factors OsWRKY62 and OsWRKY76 can undergo AS, producing truncated splice variants with a dominant-negative function; they state that this is demonstrates that there is AS-mediated feedback regulation for the WRKY transcription factor family. Investigating if any of the transcription factors that undergo *B. cinerea*-mediated DAS in wild-type lines

produce truncated proteins that could act as dominant repressors would help to determine if this feedback mechanism occurs in Arabidopsis transcription factors. If this occurs in those genes only DASed in wild-type lines it could potentially indicate that components of the MAC could be involved in modulating the feedback mechanisms that regulate transcription factors. However, considerable further work would be required to test this hypothesis.

In conclusion, this work indicates that there is a complex interconnecting splicing network involved in the regulation of the *B. cinerea* defence response, which AtCDC5 and MOS4 are part of. Both AtCDC5 and MOS4 alter the DAS of splicing factor associated genes, potentially explaining the breadth of effects observed due to DASed genes that are involved in splicing potentially having a cascade effect cumulating in a plethora of downstream genes having altered DAS. Thus the *B. cinerea* susceptibility phenotype we observe in the $Atcdc5_{KO}$ and $mos4_{KO}$ mutants is due to a combination of these multiple effects. This indicates that resistance to *B. cinerea* involves the combined effect of multiple DAS events occurring in a plethora of defence-related genes in a controlled manner.

6 Conclusion

This work shows that *B. cinerea* infection has wide ranging effects on the Arabidopsis transcriptome, with approximately one third of intron containing genes displaying *B. cinerea*-mediated DAS. In much the same way as micro-arrays highlighted the massive changes in gene expression in response to stress, RNA-Seq is doing the same for differential alternative splicing. Here we show that *B. cinerea*-mediated DAS is highly likely to be a real biological phenomenon, rather than a result of noise produced by the inherently stochastic nature of the splicing machinery and begin to elucidate some of the regulatory mechanisms.

6.1 Key defence-related genes are affected by *B. cinerea*-mediated differential alternative splicing.

We have identified a wide variety of defence-related genes that undergo *B. cinerea*-mediated DAS from those involved in perception of the organisms, i. e. *CERK1*, through to those that feed into gene expression networks such as *JAZ10*, *JAR1* and *NINJA* that are involved in regulating the JA signalling network and hence have the ability to effect JA mediated defence genes, through to transcription factors that have a more direct role in impacting the gene expression networks

6.2 *AT4G39270* links *B. cinerea*-mediated DAS with a defence phenotype.

AT4G39270 encodes a LRR-RLK that due to its structure we predict to encode a PRR. We have demonstrated that in uninfected leaves the fully spliced transcript, which is predicted to produce the LRR-RLK protein, is present in conjunction with a transcript where the first intron is retained. The retained intron transcript is likely to be retained in the nucleus and thus not produce protein. In *B. cinerea* infected, or Chitosan treated leaves the ratio of the transcripts containing the retained intron to spliced intron transcripts decrease, resulting in more functional transcript being produced. We hypothesise that this *B. cinerea*-mediated DAS would result in a rapid response to infection, because it would circumvent the time required for transcriptional activation and mRNA accumulation; the mRNA is “waiting in the nucleus” to be spliced to make proteins. We hypothesis that *AT4G39270* encodes a PRR that upon detection of a PAMP (specifically chitin) undergoes DAS, rapidly increasing protein levels, this in turn could make the plant more sensitive to detecting the pathogen resulting

in the threshold for initiation of defence responses being reached sooner. Furthermore, we have demonstrated that expression of this gene influences resistance to *B. cinerea* and that at least one of the components of the MAC affects its DAS.

6.3 *B. cinerea*-mediated differential alternative splicing enables a rapid response to infection.

B. cinerea-mediated differential alternative splicing that has the potential to alter protein levels could enable plants to respond quicker to the detection of pathogens via circumventing the time required for transcript initiation and mRNA accumulation. In addition to DASed genes that contain RIs not exiting the nucleus, two other main methods could potentially alter protein concentrations via *B. cinerea*-mediated DAS; PTC coupled to NMD and altered mRNA stability due to DAS events in the 5'UTR.

6.4 *B. cinerea*-mediated differential alternative splicing potentially regulated transcript stability via altering RNA secondary structures.

Remy *et al* 2014 determined that an AS RI event in the 5'UTR of ZIF2 results in altered secondary structure of RNA. They stated that this type of regulation, i.e. altered stability due to the DAS event affecting mRNA secondary structure was unique to their gene of interest (Remy *et al*., 2014). However, we have identified multiple cases of RI in the 5'UTRs undergoing *B. cinerea*-mediated DAS. Therefore, we hypothesise that DASed events affecting stability of transcripts via altering RNA secondary structures could be a general mechanism employed by plants to regulate protein levels. Further work at the genome-wide level to investigate if/how *B. cinerea*-mediated DAS events in the 5'UTR affects RNA secondary structures and if this impacts upon the stability of the transcript would be required to test this hypothesis.

6.5 *B. cinerea*-mediated differential alternative splicing has the potential to alter signal transduction by affecting functional domains of proteins.

Our work shows that there is a proportional change in transcripts that putatively encode truncated proteins or proteins lacking a functional domain in infected leaves compared to mock-inoculated leaves. Both of these protein modifications have the potential to alter

signal transduction; by either altering negative feedback mechanisms, as is the case of JAZ10.3, or by producing putative adapter proteins such as *EFR_ID1* that encode LRR-RLK that lack the tandem LRR due to an A5SS. Further work, investigating if both the fully spliced transcript and the transcript that encodes either truncated protein or ones with missing domains would provide additional information of how signal transduction is altered in response to pathogens.

6.6 *B. cinerea*-mediated differential alternative splicing affects networks of known defence-related genes

As mentioned earlier one of the main PRRs involved in the *B. cinerea* defence response, *CERK1*, undergoes *B. cinerea*-mediated DAS. SUA and RSN2 have been shown to be required for the proper splicing of CERK1 (Zhang *et al*., 2014), and interestingly we found that these genes are also subject to *B. cinerea*-mediated DAS, which in turn could potentially affect the DAS of *CERK1*. In addition, our motif analysis detects an enriched motif in *B. cinerea*-mediated RI events, of the human homologue of *SUA*, RBM5, thus indicating that *B. cinerea*-mediated DAS forms part of the known complex network of defence-related genes. Further investigating and expanding this network with the aim of integrating it into known gene expression networks would further expand our knowledge of the plant defence response.

6.7 Splicing factor associated genes appear to form highly interconnected networks

We have shown that a variety of splicing associated factor genes undergo *B. cinerea*-mediated differential splicing. Of particular note is the SR protein, SRZ22a, and the SCL subfamily. *SRZ22a* potentially only undergoes DAS in response to fungal and/or necrotrophic organisms. However, further work focusing on how this gene splicing patterns alter in response to infection by a variety of pathogens would be required to confirm this. In addition, we have identified that *B. cinerea*-mediated DASed RI events are enriched for motifs that putatively bind members of the SCL family, with one member of this family, *SCL30a*, having altered DAS transcript ratios in MAC mutants. Further work to investigate if the motifs that we have identified result in introns being retained in a stress specific manner

could significantly help in the development of a splicing code for plants, potentially enabling prediction of pathogen or stress related DAS events.

6.8 *AtCDC5* and *MOS4* have divergent but overlapping regulatory roles in plant defence

Our work on the MAC mutants *Atcdc5*$_{KO}$ and *mos4*$_{KO}$ highlighted the interconnectivity of the splicing factor associated genes. We have demonstrated that MAC mutants resulted in altered DAS of SR proteins, and other splicing factor associated genes, including those belonging to the MAC. This indicates that there are regulatory connections between different splicing associated factor families. We have demonstrated that perturbation of the network, i.e. by removing one of the components, has wide ranging effects on the DAS patterns of genes with multiple underlying biological functions. Specifically, we identified a subset of genes where the DAS patterns appear to be regulated by AtCDC5, or MOS4, or a combination of them both.

6.9 There is a core set of genes that undergo differentially alternatively splicing in response to stress

Comparing genes that have been identified to be DASed in response to different stresses, *B. cinerea*, salt and *Pst* indicates that there are some genes that are DASed in response to all stresses. This could indicate that there is a core set of genes that undergo stress mediated DAS, potentially with a set of stress specific DAS events feeding in. However, it was not possible to check the second part of this hypothesis, because the studies we used for comparisons had too few biological replicates to perform equivocal analyses. Validating a subset of the genes that have been identified to undergo DAS under salt, *Pst* and *B. cinerea* infection could potentially identify potential core stress genes that could be incorporated into expression networks. Identifying putative homologues of these in crop species and determining if these are also DASed could help to select potential candidate stress responsive genes that could be targets for breeding or genetic modification to improve crops resilience to stress.

6.10 Practical considerations

We have demonstrated that to compare different experiments the same analytical techniques must be employed, otherwise the difference detected may represent differences in analytical technique, rather than biological relevance. This is particularly important for comparing published DAS datasets; ideally raw data would need to be obtained (for example from the SRA archive) and DAS genes determined using the same analytical techniques for comparisons to be drawn rather than comparing published DAS datasets. Utilising *B. cinerea* marker genes identified in Chapter 3 we concluded in Chapter 4 that multiple pipelines and levels of analysis provide a more comprehensive dataset of DASed genes. However, it should be noted that this significantly increases the analytical and computational time required. In Chapter 5 we utilised only the event level, because we concluded that this had the best effort:reward ratio, considering that we were looking for an overview of how *B. cinerea*-mediated DAS changed in mutant lines. We acknowledge the fact that there are likely to be additional genes that have different *B. cinerea*-mediated DAS between wild-type and MAC mutants, but our aim was to get a general overview of the change rather than an exhaustive list of genes affected.

6.11 Concluding remarks

B. cinerea causes serious losses in more than 200 crop species worldwide (FAO, 2009; Williamson *et al*., 2007), and is estimated to cost the food industry in excess of €1,000,000,000 a year (Dean *et al*., 2012; Holmes *et al*., 2013). In addition, a growing world population combined with finite resources for growing crops, producing plants that can adapt well to stress is becoming very important to enable food to be produced in sub optimal environments. Thus understanding how plants respond to stress, and being able to use a "splicing code" to select plants with putative increased stress resilience for breeding, has the potential to have a significant beneficial impact on both economic and social implications of food security. The work carried out here has added to the body of evidence that shows that in response to stress genome-wide differential alternative splicing occurs. It has identified some putative DAS stress core genes, and splicing regulatory element motifs, which with additional work could be used to help design crops with added disease resistance contributing towards food security in the 21st century.

6.12 Future work

There are three main areas where I believe further work would be particularly useful, covering both short-term and long-term research agendas. Firstly, validation of additional defence related DASed genes, in particular those associated with CERK1 and PBL27, and integrating this information into the known expression networks would further our knowledge on how plants respond to pathogen infections. Secondly validating a subset of the genes that were identified to undergo DAS under salt, *Pst* and *B. cinerea* infection would identify potential core stress genes that could be incorporated into expression networks. Identifying putative homologues of these in crop species and determining if these also undergo stress-mediated DAS could help select potential candidate stress responsive genes that could be targets for breeding or genetic modification to improve crops resilience to stress. Lastly, further work on the identified putative defence related motifs could help to build up a splicing code. Incorporating this with theoretical work that could predict the percentage change of introns being retained under stress conditions, based on DNA sequence and predicting protein changes, could help to achieve the long-term goal of identifying plants and/or genes for breeding and crop engineering programs to help maintain food security in the 21st century.

Bibliography

Abuqamar, S., Chen, X., Dhawan, R., Bluhm, B., Salmeron, J., Lam, S., *et al*. (2006). Expression profiling and mutant analysis reveals complex regulatory networks involved in Arabidopsis response to Botrytis infection. *The Plant Journal*, *48*(1), 28–44. http://doi.org/10.1111/j.1365-313X.2006.02849.x

Alamancos, G. P., Pagès, A., Trincado, J. L., Bellora, N., & Eyras, E. (2015). Leveraging transcript quantification for fast computation of alternative splicing profiles. *RNA (New York, N.Y.)*, *21*(9), 1521–1531. http://doi.org/10.1261/rna.051557.115

Alonso, J. M., Hirayama, T., Roman, G., Nourizadeh, S., & Ecker, J. R. (1999). EIN2, a bifunctional transducer of ethylene and stress responses in Arabidopsis. *Science*, *284*(5423), 2148–2152.

Amrani, N., Dong, S., He, F., Ganesan, R., Ghosh, S., Kervestin, S., *et al*. (2006). Aberrant termination triggers nonsense-mediated mRNA decay. *Biochemical Society Transactions*, *34*(Pt 1), 39–42. http://doi.org/10.1042/BST20060039

Amrani, N., Ganesan, R., Kervestin, S., Mangus, D. A., Ghosh, S., & Jacobson, A. (2004). A faux 3'-UTR promotes aberrant termination and triggers nonsense-mediated mRNA decay. *Nature*, *7013*, 112–118.

Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data., *31*(2), 166–169. http://doi.org/10.1093/bioinformatics/btu638

Anders, S., Reyes, A., & Huber, W. (2012). Detecting differential usage of exons from RNA-seq data. *Genome Research*, *22*(10), 2008–2017. http://doi.org/10.1101/gr.133744.111

Arabidopsis Genome Initiative. (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature*, *408*(6814), 796–815. doi:10.1038/35048692

Arciga-Reyes, L., Wootton, L., Kieffer, M., & Davies, B. (2006). UPF1 is required for nonsense-mediated mRNA decay (NMD) and RNAi in Arabidopsis. *The Plant Journal : for Cell and Molecular Biology*, *47*(3), 480–489. http://doi.org/10.1111/j.1365-313X.2006.02802.x

Asai, T., Tena, G., Plotnikova, J., Willmann, M. R., Chiu, W.-L., Gomez-Gomez, L., *et al*. (2002). MAP kinase signalling cascade in Arabidopsis innate immunity. *Nature*, *415*(6875), 977–983. http://doi.org/10.1038/415977a

Aschoff, M., Hotz-Wagenblatt, A., Glatting, K.-H., Fischer, M., Eils, R., & König, R. (2013). SplicingCompass: differential splicing detection using RNA-seq data. *Bioinformatics (Oxford, England)*, *29*(9), 1141–1148. http://doi.org/10.1093/bioinformatics/btt101

Au, K. F., Jiang, H., Lin, L., Xing, Y., & Wong, W. H. (2010). Detection of splice junctions from paired-end RNA-seq data by SpliceMap. *Nucleic Acids Research*, *38*(14), 4570–4578. http://doi.org/10.1093/nar/gkq211

Bai, Y., Meng, Y., Huang, D., Qi, Y., & Chen, M. (2011). Origin and evolutionary analysis of the plant-specific TIFY transcription factor family. *Genomics*, *Aug;98*(2), 128-36. doi:10.1016/j.ygeno.2011.05.002.

Bailey, T. L. (2011). DREME: motif discovery in transcription factor ChIP-seq data. *Journal of Gerontology*, *27*(12), 1653–1659. http://doi.org/10.1093/bioinformatics/btr261

Ballut, L., Marchadier, B., Baguet, A., Tomasetto, C., Séraphin, B., & Le Hir, H. (2005). The exon junction core complex is locked onto RNA by inhibition of eIF4AIII ATPase activity. *Nature Structural & Molecular Biology*, *12*(10), 861–869. http://doi.org/10.1038/nsmb990

Barbazuk, W. B., Fu, Y., & McGinnis, K. M. (2008). Genome-wide analyses of alternative splicing in plants: opportunities and challenges. *Genome Research*, *18*(9), 1381–1392. http://doi.org/10.1101/gr.053678.106

Barta, A., Kalyna, M., & Reddy, A. S. N. (2010). Implementing a rational and consistent nomenclature for serine/arginine-rich protein splicing factors (SR proteins) in plants. *Plant Cell*, *22*(9), 2926–2929. http://doi.org/10.1105/tpc.110.078352

Bechara, E. G., Sebestyén, E., Bernardis, I., Eyras, E., & Valcárcel, J. (2013). RBM5, 6, and 10 differentially regulate NUMB alternative splicing to control cancer cell proliferation. *Molecular Cell*, *52*(5), 720–733. http://doi.org/10.1016/j.molcel.2013.11.010

Behm-Ansmant, I., Gatfield, D., Rehwinkel, J., Hilgers, V., & Izaurralde, E. (2007). A conserved role for cytoplasmic poly(A)-binding protein 1 (PABPC1) in nonsense-mediated mRNA decay. *EMBO Journal*, *26*(6), 1591–1601. http://doi.org/10.1038/sj.emboj.7601588

Bent, A. F., & Mackey, D. (2007). Elicitors, effectors, and R genes: the new paradigm and a lifetime supply of questions. *Annual Review of Phytopathology*, *45*, 399–436. http://doi.org/10.1146/annurev.phyto.45.062806.094427

Berglund, J. A., Chua, K., Abovich, N., Reed, R., & Rosbash, M. (1997). The Splicing Factor BBP Interacts Specifically with the Pre-mRNA Branchpoint Sequence UACUAAC. *Cell*, *89*(5), 781–787. http://doi.org/10.1016/S0092-8674(00)80261-5

Bhargava, A., Clabaugh, I., To, J. P., Maxwell, B. B., Chiang, Y.-H., Schaller, E. G., *et al*. (2013). Identification of Cytokinin Responsive Genes Using Microarray Meta-analysis and RNA-seq in Arabidopsis thaliana. *Plant Physiology*. 2013 May;162(1):272-94. http://doi.org/10.1104/pp.113.217026

Birkenbihl, R. P., & Somssich, I. E. (2011). Transcriptional plant responses critical for resistance towards necrotrophic pathogens. *Frontiers in Plant Science*, *2*, 76. http://doi.org/10.3389/fpls.2011.00076

Birkenbihl, R. P., Diezel, C., & Somssich, I. E. (2012). Arabidopsis WRKY33 is a key transcriptional regulator of hormonal and metabolic responses toward Botrytis cinerea infection. *Plant Physiology*, *159*(1), 266–285. http://doi.org/10.1104/pp.111.192641

Black, D. L. (2003). Mechanisms of Alternative Pre-Messenger RNA Splicing. *Annual Review of Biochemistry*, *72*(1), 291–336. http://doi.org/10.1146/annurev.biochem.72.121801.161720

Blanchette, M., & Chabot, B. (1999). Modulation of exon skipping by high-affinity hnRNP A1-binding sites and by intron elements that repress splice site utilization. *The EMBO Journal*, *18*(7), 1939–1952. http://doi.org/10.1093/emboj/18.7.1939

Blankenberg, D. D., Kuster, Von, G. G., Coraor, N. N., Ananda, G. G., Lazarus, R. R., Mangan, M. M., *et al*. (2010). Galaxy: a web-based genome analysis tool for experimentalists. *Audio, Transactions of the IRE Professional Group on*, *Chapter 19*, Unit–Un21. http://doi.org/10.1002/0471142727.mb1910s89

Blencowe, B. J. (2006). Alternative splicing: new insights from global analyses. *Cell*, *126*(1), 37–47. http://doi.org/10.1016/j.cell.2006.06.023

Brooks, A. N., Yang, L., Duff, M. O., Hansen, K. D., Park, J. W., Dudoit, S., *et al*. (2011). Conservation of an RNA regulatory map between Drosophila and mammals. *Genome Research*, *21*(2), 193–202. http://doi.org/10.1101/gr.108662.110

Brown, J. W. S., & Simpson, C. G. (1998). SPLICE SITE SELECTION IN PLANT PRE-mRNA SPLICING. *Annual Review of Plant Physiology and Plant Molecular Biology*, *49*(1), 77–95. http://doi.org/10.1146/annurev.arplant.49.1.77

Brutus, A., Sicilia, F., Macone, A., Cervone, F., & De Lorenzo, G. (2010). A domain swap approach reveals a role of the plant wall-associated kinase 1 (WAK1) as a receptor of oligogalacturonides. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(20), 9452–9457. http://doi.org/10.1073/pnas.1000675107

Bryant, D. W., Shen, R., Priest, H. D., Wong, W. K., & Mockler, T. C. (2010). Supersplat--spliced RNA-seq alignment. *Bioinformatics (Oxford, England)*, *26*(12), 1500–1505. http://doi.org/10.1093/bioinformatics/btq206

Caceres, J. F., Screaton, G. R., & Krainer, A. R. (1998). A specific subset of SR proteins shuttles continuously between the nucleus and the cytoplasm. *Genes & Development*, *12*(1), 55–66. http://doi.org/10.1101/gad.12.1.55

Campbell, M. A., Haas, B. J., Hamilton, J. P., Mount, S. M., & Buell, C. R. (2006). Comprehensive analysis of alternative splicing in rice and comparative analyses with Arabidopsis. *BMC Genomics*, *7*, 327–327. http://doi.org/10.1186/1471-2164-7-327

Cano, F., Miranda-Saavedra, D., & Lehner, P. J. (2010). RNA-binding E3 ubiquitin ligases: novel players in nucleic acid regulation. *Biochemical Society Transactions*, *38*(6), 1621–1626. http://doi.org/10.1042/BST0381621

Cao, Y., Liang, Y., Tanaka, K., Nguyen, C. T., Jedrzejczak, R. P., Joachimiak, A., & Stacey, G. (2014). The kinase LYK5 is a major chitin receptor in Arabidopsis and forms a chitin-induced complex with related kinase CERK1. *eLife*, 2014;3:e03766. http://doi.org/10.7554/eLife.03766

Chamala, S., Feng, G., Chavarro, C., & Barbazuk, W. B. (2015). Genome-wide identification of evolutionarily conserved alternative splicing events in flowering plants. *Frontiers in Bioengineering and Biotechnology*, 3, 33. http://doi.org/10.3389/fbioe.2015.00033

Chan, S.-P., & Cheng, S.-C. (2005). The Prp19-associated complex is required for specifying interactions of U5 and U6 with pre-mRNA during spliceosome activation. *Journal of Biological Chemistry*, *280*(35), 31190–31199. http://doi.org/10.1074/jbc.M505060200

Chan, S.-P., Kao, D.-I., Tsai, W.-Y., & Cheng, S.-C. (2003). The Prp19p-associated complex in spliceosome activation. *Science*, *302*(5643), 279–282. http://doi.org/10.1126/science.1086602

Chanarat, S., & Sträßer, K. (2013). Splicing and beyond: the many faces of the Prp19 complex. *Biochimica Et Biophysica Acta*, *1833*(10), 2126–2134. http://doi.org/10.1016/j.bbamcr.2013.05.023

Chang, C.-Y., Lin, W.-D., & Tu, S.-L. (2014). Genome-Wide Analysis of Heat-Sensitive Alternative Splicing in Physcomitrella patens. *Plant Physiology*, *165*(2), 826–840. http://doi.org/10.1104/pp.113.230540

Chang, Y.-F., Imam, J. S., & Wilkinson, M. F. (2007). The Nonsense-Mediated Decay RNA Surveillance Pathway. *Annual Review of Biochemistry*, *76*(1), 51–74. http://doi.org/10.1146/annurev.biochem.76.050106.093909

Chao, Q., Rothenberg, M., Solano, R., Roman, G., Terzaghi, W., & Ecker, J. R. (1997). Activation of the ethylene gas response pathway in Arabidopsis by the nuclear protein ETHYLENE-INSENSITIVE3 and related proteins. *Cell*, *89*(7), 1133–1144. http://doi.org/10.1016/S0092-8674(00)80300-1

Chen, L., Zhang, L., & Yu, D. (2010). Wounding-induced WRKY8 is involved in basal defense in Arabidopsis. *Molecular Plant-Microbe Interactions*, *23*(5), 558–565. http://doi.org/10.1094/MPMI-23-5-0558

Chen, M., & Manley, J. L. (2009). Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nature Reviews Molecular Cell Biology*, *10*(11), 741–754. http://doi.org/10.1038/nrm2777

Chini, A., Fonseca, S., Fernández, G., Adie, B., Chico, J. M., Lorenzo, O., *et al*. (2007). The JAZ family of repressors is the missing link in jasmonate signalling. *Nature*, *448*(7154), 666–671. http://doi.org/10.1038/nature06006

Cho, S., Hoang, A., Sinha, R., Zhong, X.-Y., Fu, X.-D., Krainer, A. R., & Ghosh, G. (2011). Interaction between the RNA binding domains of Ser-Arg splicing factor 1 and U1-70K snRNP protein determines early spliceosome assembly. *Pnas*, *108*(20), 8233–8238. http://doi.org/10.1073/pnas.1017700108

Chung, B. Y. W., Simons, C., Firth, A. E., Brown, C. M., & Hellens, R. P. (2006). Effect of 5'UTR introns on gene expression in Arabidopsis thaliana. *BMC Genomics*, *7*, 120. http://doi.org/10.1186/1471-2164-7-120

Chung, H. S., Cooke, T. F., DePew, C. L., Patel, L. C., Ogawa, N., Kobayashi, Y., & Howe, G. A. (2010). Alternative splicing expands the repertoire of dominant JAZ repressors of jasmonate signaling. *The Plant Journal*, *63*(4), 613–622. http://doi.org/10.1111/j.1365-313X.2010.04265.x

Colcombet, J., & Hirt, H. (2008). Arabidopsis MAPKs: a complex signalling network involved in multiple biological processes. *Biochemical Journal*, *413*(2), 217–226. http://doi.org/10.1042/BJ20080625

Cooke, E. (2013). Transcriptional analysis of the interaction between Botrytis cinerea and a host Arabidopsis thaliana using high-throughput data. *University of Warwick*, *Phd thesis*, 1–257.

Culler, S. J., Hoff, K. G., Voelker, R. B., Berglund, J. A., & Smolke, C. D. (2010). Functional selection and systematic analysis of intronic splicing elements identify active sequence motifs and associated splicing factors. *Nucleic Acids Research*, *38*(15), 5152–5165. http://doi.org/10.1093/nar/gkq248

Dangl, J. L., & Jones, J. D. (2001). Plant pathogens and integrated defence responses to infection. *Nature*, *411*(6839), 826–833. http://doi.org/10.1038/35081161

De Bona, F. F., Ossowski, S. S., Schneeberger, K. K., & Rätsch, G. G. (2008). Optimal spliced alignments of short sequence reads. *Audio, Transactions of the IRE Professional Group on*, *24*(16), i174–i180. http://doi.org/10.1093/bioinformatics/btn300

De Conti, L., Baralle, M., & Buratti, E. (2012). Exon and intron definition in pre-mRNA splicing. *Wiley Interdisciplinary Reviews: RNA*, *4*(1), 49–60. http://doi.org/10.1002/wrna.1140

de Lima Morais, D. A., & Harrison, P. M. (2010). Large-scale evidence for conservation of NMD candidature across mammals. *PloS One*, *5*(7), e11695. http://doi.org/10.1371/journal.pone.0011695

Dean, R., van Kan, J. A. L., Pretorius, Z. A., Hammond-Kosack, K. E., Di Pietro, A., Spanu, P. D., *et al*. (2012). The Top 10 fungal pathogens in molecular plant pathology. *Molecular Plant Pathology*, *13*(4), 414–430. http://doi.org/10.1111/j.1364-3703.2011.00783.x

Delaney, T. P., Uknes, S., Vernooij, B., Friedrich, L., Weymann, K., Negrotto, D., *et al*. (1994). A central role of salicylic Acid in plant disease resistance. *Science*, *266*(5188), 1247–1250. http://doi.org/10.1126/science.266.5188.1247

Denby, K. J., Kumar, P., & Kliebenstein, D. J. (2004). Identification of Botrytis cinerea susceptibility loci in Arabidopsis thaliana. *The Plant Journal*, *38*(3), 473–486. http://doi.org/10.1111/j.0960-7412.2004.02059.x

Deutsch, M., & Long, M. (1999). Intron-exon structures of eukaryotic model organisms. *Nucleic Acids Research*, *27*(15), 3219–3228.

Devoto, A., Nieto-Rostro, M., Xie, D., Ellis, C., Harmston, R., Patrick, E., *et al*. (2002). COI1 links jasmonate signalling and fertility to the SCF ubiquitin-ligase complex in Arabidopsis. *The Plant Journal* , *32*(4), 457–466. http://dx.doi.org/10.1046/j.1365-313X.2002.01432.x

Dinesh-Kumar, S. P., & Baker, B. J. (2000). Alternatively spliced N resistance gene transcripts: their possible role in tobacco mosaic virus resistance. *PNAS*, *97*(4), 1908–1913. http://doi.org/10.2307/121575

Ding, F., Cui, P., Wang, Z., Zhang, S., Ali, S., & Xiong, L. (2014). Genome-wide analysis of alternative splicing of pre-mRNA under salt stress in Arabidopsis. *BMC Genomics*, *15*, 431. http://doi.org/10.1186/1471-2164-15-431

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., *et al*. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)*, *29*(1), 15–21. http://doi.org/10.1093/bioinformatics/bts635

Dogan, R. I., Getoor, L., Wilbur, W. J., & Mount, S. M. (2007). SplicePort--an interactive splice-site analysis tool. *Nucleic Acids Research*, *35*(Web Server issue), W285–W291. http://doi.org/10.1093/nar/gkm407

Drechsel, G., Kahles, A., Kesarwani, A. K., Stauffer, E., Behr, J., Drewe, P., *et al*. (2013). Nonsense-mediated decay of alternative precursor mRNA splicing variants is a major determinant of the Arabidopsis steady state transcriptome. *The Plant Cell*, *25*(10), 3726–3742. http://doi.org/10.1105/tpc.113.115485

Duque, P. (2011). A role for SR proteins in plant stress responses. *Plant Signaling & Behavior*, *6*(1), 49–54. http://doi.org/10.4161/psb.6.1.14063

Engström, P. G., Steijger, T., Sipos, B., Grant, G. R., Kahles, A., Rätsch, G., *et al*. (2013). Systematic evaluation of spliced alignment programs for RNA-seq data. *Nature Methods*, *10*(12), 1185–1191. http://doi.org/10.1038/nmeth.2722

Eulgem, T. (2006). Dissecting the WRKY web of plant defense regulators. *PLoS Pathogens*, *2*(11), e126. http://doi.org/10.1371/journal.ppat.0020126

Fairbrother, W. G., Yeo, G. W., Yeh, R., Goldstein, P., Mawson, M., Sharp, P. A., & Burge, C. B. (2004). RESCUE-ESE identifies candidate exonic splicing enhancers in vertebrate exons. *Nucleic Acids Research*, *32*(Web Server issue), W187–90. http://doi.org/10.1093/nar/gkh393

FAO. (2009). How to Feed the World in 2050 – Global Agriculture Towards 2050. FAO, Rome.

Feldmann, K. A. (1991). T-DNA insertion mutagenesis in Arabidopsis: mutational spectrum *The Plant* Journal, 1: 71–82. doi:10.1111/j.1365-313X.1991.00071.x

Ferrari, S., Galletti, R., Denoux, C., De Lorenzo, G., Ausubel, F. M., & Dewdney, J. (2007). Resistance to Botrytis cinerea Induced in Arabidopsis by Elicitors Is Independent of Salicylic Acid, Ethylene, or Jasmonate Signaling But Requires PHYTOALEXIN DEFICIENT3. *Plant Physiology.* 144(1):367-79. DOI:10.1104/pp.107.095596

Ferrari, S., Plotnikova, J. M., De Lorenzo, G., & Ausubel, F. M. (2003). Arabidopsis local resistance to Botrytis cinerea involves salicylic acid and camalexin and requires EDS4 and PAD2, but not SID2, EDS5 or PAD4. *The Plant Journal : for Cell and Molecular Biology*, *35*(2), 193–205. http://doi.org/10.1046/j.1365-313X.2003.01794.x

Filichkin, S. A., Priest, H. D., Givan, S. A., Shen, R., Bryant, D. W., Fox, S. E., *et al*. (2010). Genome-wide mapping of alternative splicing in Arabidopsis thaliana. *Genome research*, 20(1):45-58. doi: 10.1101/gr.093302.109.

Foissac, S., & Sammeth, M. (2007). ASTALAVISTA: dynamic and flexible analysis of alternative splicing events in custom gene datasets. *Nucleic Acids Research*, *35*(Web Server), W297–W299. http://doi.org/10.1093/nar/gkm311

Fonseca, N. A., Marioni, J., & Brazma, A. (2014). RNA-Seq gene profiling--a systematic empirical comparison. *PloS One*, *9*(9), e107026–e107026. http://doi.org/10.1371/journal.pone.0107026

Foster, L. (2012, June 11). Regulation of gene expression during plant defence by alternative splicing (Master's thesis). (K. J. Denby, Ed.). The University of Warwick.

Frilander, M. J., & Steitz, J. A. (1999). Initial recognition of U12-dependent introns requires both U11/5. *Genes & Development*, 13(7), 851–863.

Fu, Z. Q., Guo, M., Jeong, B.-R., Tian, F., Elthon, T. E., Cerny, R. L., *et al*. (2007). A type III effector ADP-ribosylates RNA-binding proteins and quells plant immunity. *Nature*, *447*(7142), 284–288. http://doi.org/10.1038/nature05737

Fujita, M., Fujita, Y., Noutoshi, Y., Takahashi, F., Narusaka, Y., Yamaguchi-Shinozaki, K., & Shinozaki, K. (2006). Crosstalk between abiotic and biotic stress responses: a current view from the points of convergence in the stress signaling networks. *Current Opinion in Plant Biology*, *9*(4), 436–442. http://doi.org/10.1016/j.pbi.2006.05.014

Garber, M., Grabherr, M. G., Guttman, M., & Trapnell, C. (2011). Computational methods for transcriptome annotation and quantification using RNA-seq. *Nature Methods*, *8*(6), 469–477. http://doi.org/10.1038/nmeth.1613

Gassmann, W. (2008). Alternative Splicing in Plant Defense. *Current Topics in Microbiology and Immunology*, (326), 219–233.

Gatto, A., Torroja-Fungairiño, C., Mazzarotto, F., Cook, S. A., Barton, P. J. R., Sánchez-Cabo, F., & Lara-Pezzi, E. (2014). FineSplice, enhanced splice junction detection and quantification: a novel pipeline based on the assessment of diverse RNA-Seq alignment solutions. *Nucleic Acids Research*, *42*(8), e71–e71. http://doi.org/10.1093/nar/gku166

Giardine, B., Riemer, C., Hardison, R. C., Burhans, R., Elnitski, L., Shah, P., *et al*. (2005). Galaxy: a platform for interactive large-scale genome analysis. *Genome Research*, *15*(10), 1451–1455. http://doi.org/10.1101/gr.4086505

Glazebrook, J. (2005). Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. *Annual Review of Phytopathology*, *43*, 205–227. http://doi.org/10.1146/annurev.phyto.43.040204.135923

Gloggnitzer, J., Akimcheva, S., Srinivasan, A., Kusenda, B., Riehs, N., Stampfl, H., *et al*. (2014). Nonsense-mediated mRNA decay modulates immune receptor levels to regulate plant antibacterial defense. *Cell Host & Microbe*, *16*(3), 376–390. http://doi.org/10.1016/j.chom.2014.08.010

Goecks, J., Nekrutenko, A., Taylor, J., & Team, T. G. (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol*, *11*(8), R86–R86. http://doi.org/10.1186/gb-2010-11-8-r86

Goff, S. A., Vaughn, M., McKay, S., Lyons, E., Stapleton, A. E., Gessler, D., *et al*. (2011). The iPlant Collaborative: Cyberinfrastructure for Plant Biology. *Frontiers in Plant Science*, *2*, 34–34. http://doi.org/10.3389/fpls.2011.00034

Golovkin, M., & Reddy, A. S. (1999). An SC35-like protein and a novel serine/arginine-rich protein interact with Arabidopsis U1-70K protein. *Journal of Biological Chemistry*, *274*(51), 36428–36438.

Goodall, G. J., & Filipowicz, W. (1989). The AU-rich sequences present in the introns of plant nuclear pre-mRNAs are required for splicing. *Cell*, *58*(3), 473–483.

Gosai, S. J., Foley, S. W., Wang, D., Silverman, I. M., Selamoglu, N., Nelson, A. D. L., *et al*. (2015). Global analysis of the RNA-protein interaction and RNA secondary structure landscapes of the Arabidopsis nucleus. *Molecular Cell*, *57*(2), 376–388. http://doi.org/10.1016/j.molcel.2014.12.004

Gou, X., He, K., Yang, H., Yuan, T., Lin, H., Clouse, S. D., & Li, J. (2010). Genome-wide cloning and sequence analysis of leucine-rich repeat receptor-like protein kinase genes in Arabidopsis thaliana. *BMC Genomics*, *11*, 19. http://doi.org/10.1186/1471-2164-11-19

Göhring, J., Jacak, J., & Barta, A. (2014). Imaging of endogenous messenger RNA splice variants in living cells reveals nuclear retention of transcripts inaccessible to nonsense-mediated decay in Arabidopsis. *The Plant Cell*, *26*(2), 754–764. http://doi.org/10.1105/tpc.113.118075

Grant, G. R., Farkas, M. H., Pizarro, A. D., Lahens, N. F., Schug, J., Brunk, B. P., *et al*. (2011). Comparative analysis of RNA-Seq alignment algorithms and the RNA-Seq unified mapper (RUM). *Bioinformatics (Oxford, England)*, *27*(18), 2518–2528. http://doi.org/10.1093/bioinformatics/btr427

Greeff, C., Roux, M., Mundy, J., & Petersen, M. (2012). Receptor-like kinase complexes in plant innate immunity. *Frontiers in Plant Science*, *3*, 209–209. http://doi.org/10.3389/fpls.2012.00209

Griffith, M., Griffith, O. L., Mwenifumbo, J., Goya, R., Morrissy, A. S., Morin, R. D., *et al*. (2010). Alternative expression analysis by RNA sequencing. *Nature Methods*, *7*(10), 843–847. http://doi.org/10.1038/nmeth.1503

Gualandi, F., Rimessi, P., Cardazzo, B., Toffolatti, L., Dunckley, M. G., Calzolari, E., *et al*. (2003). Genomic definition of a pure intronic dystrophin deletion responsible for an XLDC splicing mutation: in vitro mimicking and antisense modulation of the splicing abnormality. *Gene*, *311*, 25–33. http://doi.org/10.1016/S0378-1119(03)00527-4

Gulledge, A. A., Roberts, A. D., Vora, H., Patel, K., & Loraine, A. E. (2012). Mining Arabidopsis thaliana RNA-seq data with Integrated Genome Browser reveals stress-induced alternative splicing of the putative splicing regulator SR45a. *American Journal of Botany*, *99*(2), 219–231. http://doi.org/10.3732/ajb.1100355

Guo, Y., Zhao, S., Ye, F., Sheng, Q., & Shyr, Y. (2014). MultiRankSeq: multiperspective approach for RNAseq differential expression analysis and quality control. *BioMed Research International*, *2014*, 248090. http://doi.org/10.1155/2014/248090

Han, L., Li, G.-J., Yang, K.-Y., Mao, G., Wang, R., Liu, Y., & Zhang, S. (2010). Mitogen-activated protein kinase 3 and 6 regulate Botrytis cinerea-induced ethylene production in Arabidopsis. *The Plant Journal*, *64*(1), 114–127. http://doi.org/10.1111/j.1365-313X.2010.04318.x

Hayafune, M., Berisio, R., Marchetti, R., Silipo, A., Kayama, M., Desaki, Y., *et al*. (2014). Chitin-induced activation of immune signaling by the rice receptor CEBiP relies on a unique sandwich-type dimerization. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(3), E404–13. http://doi.org/10.1073/pnas.1312099111

Holmes, M., Hughes, R., Jones, G., Sturman, V., Whiting, M., Wiltshire, J., & Harden, C. (2013). A 2020 vision for the global food system. *WWF Report UK*, 1–44.

Hori, K., & Watanabe, Y. (2005). UPF3 suppresses aberrant spliced mRNA in Arabidopsis. *The Plant Journal : for Cell and Molecular Biology*, *43*(4), 530–540. http://doi.org/10.1111/j.1365-313X.2005.02473.x

Hori, K., & Watanabe, Y. (2007). Context analysis of termination codons in mRNA that are recognized by plant NMD. *Plant and Cell Physiology*, *48*(7), 1072–1078. http://doi.org/10.1093/pcp/pcm075

Howard, B. E., Hu, Q., Babaoglu, A. C., Chandra, M., Borghi, M., Tan, X., *et al*. (2013). High-throughput RNA sequencing of pseudomonas-infected Arabidopsis reveals hidden transcriptome complexity and novel splice variants. *PloS One*, *8*(10), e74183. http://doi.org/10.1371/journal.pone.0074183

Huang, S., Zhang, J., Li, R., Zhang, W., He, Z., Lam, T.-W., *et al*. (2011). SOAPsplice: Genome-Wide ab initio Detection of Splice Junctions from RNA-Seq Data. *Frontiers in Genetics*, *2*, 46–46. http://doi.org/10.3389/fgene.2011.00046

Hulsen, T., de Vlieg, J., & Alkema, W. (2008). BioVenn - a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC Genomics*, *9*, 488–488. http://doi.org/10.1186/1471-2164-9-488

Iida, K., Fukami-Kobayashi, K., Toyoda, A., Sakaki, Y., Kobayashi, M., Seki, M., & Shinozaki, K. (2009). Analysis of Multiple Occurrences of Alternative Splicing Events in Arabidopsis thaliana Using Novel Sequenced Full-Length cDNAs. *DNA Research*, *16*(3), 155–164. http://doi.org/10.1093/dnares/dsp009

Isshiki, M., Tsumoto, A., & Shimamoto, K. (2006). The serine/arginine-rich protein family in rice plays important roles in constitutive and alternative splicing of pre-mRNA. *Plant Cell*, *18*(1), 146–158. http://doi.org/10.1105/tpc.105.037069

Jens Staal, C. D. (2008). RLM3, a potential adaptor between specific TIR-NB-LRR receptors and DZC proteins. *Communicative & Integrative Biology*, *1*(1), 59.

Jeong, H.-J., Kim, Y. J., Kim, S. H., Kim, Y.-H., Lee, I.-J., Kim, Y. K., & Shin, J. S. (2011). Nonsense-Mediated mRNA Decay Factors, UPF1 and UPF3, Contribute to Plant Defense. *Plant and Cell Physiology* Dec;52(12):2147-56. doi: 10.1093/pcp/pcr144.

Kaku, H., Nishizawa, Y., Ishii-Minami, N., Akimoto-Tomiyama, C., Dohmae, N., Takio, K., *et al*. (2006). Plant cells recognize chitin fragments for defense signaling through a plasma membrane receptor. *Pnas*, *103*(29), 11086–11091. http://doi.org/10.1073/pnas.0508882103

Kalyna, M., & Barta, A. (2004). A plethora of plant serine/arginine-rich proteins: redundancy or evolution of novel gene functions? *Biochemical Society Transactions*, *32*(Pt 4), 561–564. http://doi.org/10.1042/BST0320561

Kalyna, M., Lopato, S., & Barta, A. (2003). Ectopic expression of atRSZ33 reveals its function in splicing and causes pleiotropic changes in development. *Molecular Biology of the Cell*, *14*(9), 3565–3577. http://doi.org/10.1091/mbc.E03-02-0109

Kalyna, M., Simpson, C. G., Syed, N. H., Lewandowska, D., Marquez, Y., Kusenda, B., *et al*. (2012). Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in Arabidopsis. *Nucleic Acids Research*, *40*(6), 2454–2469. http://doi.org/10.1093/nar/gkr932

Kanitz, A., Gypas, F., Gruber, A. J., Gruber, A. R., Martin, G., & Zavolan, M. (2015). Comparative assessment of methods for the computational inference of transcript isoform abundance from RNA-seq data. *Genome Biology*, *16*, 150–150. http://doi.org/10.1186/s13059-015-0702-5

Kashima, I., Yamashita, A., Izumi, N., Kataoka, N., Morishita, R., Hoshino, S., *et al*. (2006). Binding of a novel SMG-1-Upf1-eRF1-eRF3 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and nonsense-mediated mRNA decay. *Genes & Development*, *20*(3), 355–367. http://doi.org/10.1101/gad.1389006

Katz, Y., Wang, E. T., Airoldi, E. M., & Burge, C. B. (2010). Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nature Methods*, *7*(12), 1009–1015. http://doi.org/10.1038/nmeth.1528

Kennedy, C. F., Krämer, A., & Berget, S. M. (1998). A role for SRp54 during intron bridging of small introns with pyrimidine tracts upstream of the branch point. *Molecular and Cellular Biology*, *18*(9), 5425–5434.

Keren, H., Lev-Maor, G., & Ast, G. (2010). Alternative splicing and evolution: diversification, exon definition and function. *Nature Publishing Group*, *11*(5), 345–355. http://doi.org/10.1038/nrg2776

Kerényi, Z., Mérai, Z., Hiripi, L., Benkovics, A., Gyula, P., Lacomme, C., *et al*. (2008). Inter-kingdom conservation of mechanism of nonsense-mediated mRNA decay. *EMBO Journal*, *27*(11), 1585–1595. http://doi.org/10.1038/emboj.2008.88

Kertész, S., Kerényi, Z., Mérai, Z., Bartos, I., Pálfy, T., Barta, E., & Silhavy, D. (2006). Both introns and long 3'-UTRs operate as cis-acting elements to trigger nonsense-mediated decay in plants. *Nucleic Acids Research*, *34*(21), 6147–6157. http://doi.org/10.1093/nar/gkl737

Kieber, J. J., Rothenberg, M., Roman, G., Feldmann, K. A., & Ecker, J. R. (1993). CTR1, a negative regulator of the ethylene response pathway in Arabidopsis, encodes a member of the raf family of protein kinases. *Cell*, *72*(3), 427–441. http://doi.org/10.1016/0092-8674(93)90119-B

Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nature Methods*, *12*(4), 357–360. http://doi.org/10.1038/nmeth.3317

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., & Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology*, *14*(4), R36–R36. http://doi.org/10.1186/gb-2013-14-4-r36

Kim, S. H., Kwon, Il, S., Saha, D., Anyanwu, N. C., & Gassmann, W. (2009). Resistance to the Pseudomonas syringae effector HopA1 is governed by the TIR-NBS-LRR protein RPS6 and is enhanced by mutations in SRFR1. *Plant Physiology*, *150*(4), 1723–1732. http://doi.org/10.1104/pp.109.139238

Kohorn, B. D., Kohorn, S. L., Todorova, T., Baptiste, G., Stansky, K., & McCullough, M. (2012). A dominant allele of Arabidopsis pectin-binding wall-associated kinase induces a stress response suppressed by MPK6 but not MPK3 mutations. *Molecular Plant*, *5*(4), 841–851. http://doi.org/10.1093/mp/ssr096

Koncz, C., Dejong, F., Villacorta, N., Szakonyi, D., & Koncz, Z. (2012). The spliceosome-activating complex: molecular mechanisms underlying the function of a pleiotropic regulator. *Frontiers in Plant Science*, *3*, 1–12. http://doi.org/10.3389/fpls.2012.00009/abstract

Kong, Q., Qu, N., Gao, M., Zhang, Z., Ding, X., Yang, F., *et al*. (2012). The MEKK1-MKK1/MKK2-MPK4 kinase cascade negatively regulates immunity mediated by a mitogen-activated protein kinase kinase kinase in Arabidopsis. *Plant Cell*, *24*(5), 2225–2236. http://doi.org/10.1105/tpc.112.097253

Krol, E., Mentzel, T., Chinchilla, D., Boller, T., Felix, G., Kemmerling, B., *et al*. (2010). Perception of the Arabidopsis danger signal peptide 1 involves the pattern recognition receptor AtPEPR1 and its close homologue AtPEPR2. *Journal of Biological Chemistry*, *285*(18), 13471–13479. http://doi.org/10.1074/jbc.M109.097394

Kumar, K. R. R., & Kirti, P. B. (2012). Novel role for a serine/arginine-rich splicing factor, AdRSZ21 in plant defense and HR-like cell death. *Plant Molecular Biology*, *80*(4-5), 461–476. http://doi.org/10.1007/s11103-012-9960-8

Kurosaki, T., & Maquat, L. E. (2013). Rules that govern UPF1 binding to mRNA 3′ UTRs. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(9), 3357–3362. http://doi.org/10.1073/pnas.1219908110

Lai, Z., Vinod, K., Zheng, Z., Fan, B., & Chen, Z. (2008). Roles of Arabidopsis WRKY3 and WRKY4 transcription factors in plant responses to pathogens. *BMC Plant Biology*, *8*, 68–68. http://doi.org/10.1186/1471-2229-8-68

Lai, Z., Wang, F., Zheng, Z., Fan, B., & Chen, Z. (2011). A critical role of autophagy in plant resistance to necrotrophic fungal pathogens. *The Plant Journal : for Cell and Molecular Biology*, *66*(6), 953–968. http://doi.org/10.1111/j.1365-313X.2011.04553.x

Law, C. W., Chen, Y., Shi, W., & Smyth, G. K. (2014). voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biology*, *15*(2), R29–R29. http://doi.org/10.1186/gb-2014-15-2-r29

Lazar, G., & Goodman, H. M. (2000). The Arabidopsis splicing factor SR1 is regulated by alternative splicing. *Plant Molecular Biology*, *42*(4), 571–581. http://doi.org/10.1023/A:1006394207479

Le Hir, H., Gatfield, D., Izaurralde, E., & Moore, M. J. (2001). The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. *EMBO Journal*, *20*(17), 4987–4997. http://doi.org/10.1093/emboj/20.17.4987

Le Hir, H., Izaurralde, E., Maquat, L. E., & Moore, M. J. (2000). The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions. *EMBO Journal*, *19*(24), 6860–6869. http://doi.org/10.1093/emboj/19.24.6860

Le Roux, C., Del Prete, S., Boutet-Mercey, S., Perreau, F., Balagué, C., Roby, D., *et al*. (2014). The hnRNP-Q protein LIF2 participates in the plant immune response. *PloS One*, *9*(6), e99343–e99343. http://doi.org/10.1371/journal.pone.0099343

Le, M. H., Cao, Y., Zhang, X.-C., & Stacey, G. (2014). LIK1, a CERK1-interacting kinase, regulates plant immune responses in Arabidopsis. *PloS One*, *9*(7), e102245–e102245. http://doi.org/10.1371/journal.pone.0102245

Li, B., & Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, *12*, 323–323. http://doi.org/10.1186/1471-2105-12-323

Liao, Y., Smyth, G. K., & Shi, W. (2013). The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Research*. http://doi.org/10.1093/nar/gkt214
Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features., *30*(7), 923–930. http://doi.org/10.1093/bioinformatics/btt656

Lin, C. H., & Patton, J. G. (1995). Regulation of alternative 3’ splice site selection by constitutive splicing factors. *RNA (New York, N.Y.)*, *1*(3), 234–245.

Liu, J., Chen, X., Liang, X., Zhou, X., Yang, F., Liu, J., *et al*. (2016). Alternative splicing of rice WRKY62 and WRKY76 transcription factor genes in pathogen defense. *Plant Physiology*, pp.01921.2015. http://doi.org/10.1104/pp.15.01921

Liu, R., Holik, A. Z., Su, S., Jansz, N., Chen, K., Leong, H. S., *et al*. (2015a). Why weight? Modelling sample and observational level variability improves power in RNA-seq analyses. *Nucleic Acids Research*, *43*(15), e97–e97. http://doi.org/10.1093/nar/gkv412

Liu, R., Loraine, A. E., & Dickerson, J. A. (2014). Comparisons of computational methods for differential alternative splicing detection using RNA-seq in plant systems. *BMC Bioinformatics*, *15*(1), 364. http://doi.org/10.1186/s12859-014-0364-4

Liu, S., Kracher, B., Ziegler, J., Birkenbihl, R. P., & Somssich, I. E. (2015b). Negative regulation of ABA signaling by WRKY33 is critical for Arabidopsis immunity towards Botrytis cinerea 2100. *eLife*, *4*, e07295–e07295. http://doi.org/10.7554/eLife.07295

Liu, T., Liu, Z., Song, C., Hu, Y., Han, Z., She, J., *et al*. (2012). Chitin-Induced Dimerization Activates a Plant Immune Receptor. *Science*, *336*(6085), 1160–1164. http://doi.org/10.1126/science.1218867

Long, J. C., & Caceres, J. F. (2009). The SR protein family of splicing factors: master regulators of gene expression. *Biochemical Journal*, *417*(1), 15–27. http://doi.org/10.1042/BJ20081501

Longman, D., Johnstone, I. L., & Caceres, J. F. (2000). Functional characterization of SR and SR-related genes in Caenorhabditis elegans. *EMBO Journal*, *19*(7), 1625–1637. http://doi.org/10.1093/emboj/19.7.1625

Lopato, S., Kalyna, M., Dorner, S., Kobayashi, R., Krainer, A. R., & Barta, A. (1999). atSR30, one of two SF2/ASF-like proteins from *Arabidopsis thaliana,* regulates splicing of specific plant genes. *Genes and Development,* 13: 987-1001.

Lorković, Z. J., Lopato, S., Pexa, M., Lehner, R., & Barta, A. (2004). Interactions of Arabidopsis RS Domain Containing Cyclophilins with SR Proteins and U1 and U11 Small Nuclear

Ribonucleoprotein-specific Proteins Suggest Their Involvement in Pre-mRNA Splicing. *Journal of Biological Chemistry*, *279*(32), 33890–33898. http://doi.org/10.1074/jbc.M400270200

Lorković , Z. J., Wieczorek, D. A., Lambermon, M. H. L., & Filipowicz, W. (2000). Pre-mRNA splicing in higher plants. *Trends in Plant Science*, *5*(4), 160–167.

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550–550. http://doi.org/10.1186/s13059-014-0550-8

Maere, S., Heymans, K., & Kuiper, M. (2005). BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics (Oxford, England)*, *21*(16), 3448–3449. http://doi.org/10.1093/bioinformatics/bti551

Magis, A., Funk, C., & Price, N. (2015). SNAPR: a bioinformatics pipeline for efficient and accurate RNA- seq alignment and analysis. Life Sciences Letters, IEEE, *1*(2), 22–25. DOI: 10.1109/LLS.2015.2465870

Makarov, E. M., Makarova, O. V., Urlaub, H., Gentzel, M., Will, C. L., Wilm, M., & Lührmann, R. (2002). Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. *Science*, *298*(5601), 2205–2208. http://doi.org/10.1126/science.1077783

Manley, J. L., & Krainer, A. R. (2010). A rational nomenclature for serine/arginine-rich protein splicing factors (SR proteins). *Genes & Development*, *24*(11), 1073–1074. http://doi.org/10.1101/gad.1934910

Mao, G., Meng, X., Liu, Y., Zheng, Z., Chen, Z., & Zhang, S. (2011). Phosphorylation of a WRKY transcription factor by two pathogen-responsive MAPKs drives phytoalexin biosynthesis in Arabidopsis. *The Plant Cell*, *23*(4), 1639–1653. http://doi.org/10.1105/tpc.111.084996

Mao, R., Raj Kumar, P. K., Guo, C., Zhang, Y., & Liang, C. (2014). Comparative analyses between retained introns and constitutively spliced introns in Arabidopsis thaliana using

random forest and support vector machine. *PloS One*, *9*(8), e104049. http://doi.org/10.1371/journal.pone.0104049

Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., *et al*. (2015). CDD: NCBI's conserved domain database. *Nucleic Acids Research*, *43*(Database issue), D222–6. http://doi.org/10.1093/nar/gku1221

Marone, D., Russo, M. A., Laidò, G., De Leonardis, A. M., & Mastrangelo, A. M. (2013). Plant Nucleotide Binding Site-Leucine-Rich Repeat (NBS-LRR) Genes: Active Guardians in Host Defense Responses. *International Journal of Molecular Sciences*, *14*(4), 7302–7326. http://doi.org/10.3390/ijms14047302

Marquez, Y., Brown, J. W. S., Simpson, C. G., Barta, A., & Kalyna, M. (2012). Transcriptome survey reveals increased complexity of the alternative splicing landscape in Arabidopsis. *Genome Research*, *22*(6), 1184–1195. http://doi.org/10.1101/gr.134106.111

Mastrangelo, A. M., Marone, D., Laidò, G., De Leonardis, A. M., & De Vita, P. (2012). Alternative splicing: enhancing ability to cope with stress via transcriptome plasticity. *Plant Science : an International Journal of Experimental Plant Biology*, *185-186*, 40–49. http://doi.org/10.1016/j.plantsci.2011.09.006

Matsukura, S., Mizoi, J., Yoshida, T., Todaka, D., Ito, Y., Maruyama, K., *et al*. (2010). Comprehensive analysis of rice DREB2-type genes that encode transcription factors involved in the expression of abiotic stress-responsive genes. *Molecular Genetics and Genomics*, *283*(2), 185–196. http://doi.org/10.1007/s00438-009-0506-y

Mbengue, M., Navaud, O., Peyraud, R., Barascud, M., Badet, T., Vincent, R., *et al*. (2016). Emerging Trends in Molecular Interactions between Plants and the Broad Host Range Fungal Pathogens Botrytis cinerea and Sclerotinia sclerotiorum. *Frontiers in Plant Science*, *7*(15140), 987. http://doi.org/10.1105/tpc.114.128611

McGuire, A. M., Pearson, M. D., & Neafsey, D. E. (2008). Cross-kingdom patterns of alternative splicing and splice recognition. *Genome Biology,* 9:R50 DOI: 10.1186/gb-2008-9-3-r50

Melamud, E., & Moult, J. (2009). Stochastic noise in splicing machinery. *Nucleic Acids Research*. 37(14):4873-86. doi: 10.1093/nar/gkp471.

Meng, X., Xu, J., He, Y., Yang, K.-Y., Mordorski, B., Liu, Y., & Zhang, S. (2013). Phosphorylation of an ERF transcription factor by Arabidopsis MPK3/MPK6 regulates plant defense gene induction and fungal resistance. *The Plant Cell*, *25*(3), 1126–1142. http://doi.org/10.1105/tpc.112.109074

Mengiste, T., Chen, X., Salmeron, J., & Dietrich, R. (2003). The BOTRYTIS SUSCEPTIBLE1 gene encodes an R2R3MYB transcription factor protein that is required for biotic and abiotic stress responses in Arabidopsis. *The Plant Cell*, *15*(11), 2551–2565. http://doi.org/10.1105/tpc.014167

Meyers, B. C., Morgante, M., & Michelmore, R. W. (2002). TIR-X and TIR-NBS proteins: two new families related to disease resistance TIR-NBS-LRR proteins encoded in Arabidopsis and other plant genomes. *The Plant Journal*, *32*(1), 77–92. http://doi.org/10.1046/j.1365-313X.2002.01404.x

Michael Weaver, L., Swiderski, M. R., Li, Y., & Jones, J. D. G. (2006). The Arabidopsis thaliana TIR-NB-LRR R-protein, RPP1A; protein localization and constitutive activation of defence by truncated alleles in tobacco and Arabidopsis. *The Plant Journal : for Cell and Molecular Biology*, *47*(6), 829–840. http://doi.org/10.1111/j.1365-313X.2006.02834.x

Miya, A., Albert, P., Shinya, T., Desaki, Y., Ichimura, K., Shirasu, K., *et al*. (2007). CERK1, a LysM receptor kinase, is essential for chitin elicitor signaling in Arabidopsis. *Pnas*, *104*(49), 19613–19618. http://doi.org/10.1073/pnas.0705147104

Moffat, C. S., Ingle, R. A., Wathugala, D. L., Saunders, N. J., Knight, H., & Knight, M. R. (2012). ERF5 and ERF6 play redundant roles as positive regulators of JA/Et-mediated defense against Botrytis cinerea in Arabidopsis. *PloS One*, *7*(4), e35995–e35995. http://doi.org/10.1371/journal.pone.0035995

Monaghan, J., Xu, F., Gao, M., Zhao, Q., Palma, K., Long, C., *et al*. (2009). Two Prp19-like U-box proteins in the MOS4-associated complex play redundant roles in plant innate immunity. *PLoS Pathogens*, *5*(7), e1000526. http://doi.org/10.1371/journal.ppat.1000526

Monaghan, J., Xu, F., Xu, S., Zhang, Y., & Li, X. (2010). Two putative RNA-binding proteins function with unequal genetic redundancy in the MOS4-associated complex. *Plant Physiology*, *154*(4), 1783–1793. http://doi.org/10.1104/pp.110.158931

Moreno, J. E., Shyu, C., Campos, M. L., Patel, L. C., Chung, H. S., Yao, J., *et al*. (2013). Negative Feedback Control of Jasmonate Signaling by an Alternative Splice Variant of JAZ10. *Plant Physiology*, *162*(2), 1006–1017. http://doi.org/10.1104/pp.113.218164

Nagy, E., & Maquat, L. E. (1998). A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. *Trends in Biochemical Sciences*, *23*(6), 198–199.

Narusaka, Y., Shinya, T., Narusaka, M., Motoyama, N., Shimada, H., Murakami, K., & Shibuya, N. (2013). Presence of LYM2 dependent but CERK1 independent disease resistance in Arabidopsis. *Plant Signaling & Behavior*, *8*(9). http://doi.org/10.4161/psb.25345

Navarro, L., Zipfel, C., Rowland, O., Keller, I., Robatzek, S., Boller, T., & Jones, J. D. G. (2004). The Transcriptional Innate Immune Response to flg22. Interplay and Overlap with Avr Gene-Dependent Defense Responses and Bacterial Pathogenesis. *Journal of Plant Physiology*, *135*, 1113–1128.

Nelson, K. K., & Green, M. R. (1989). Mammalian U2 snRNP has a sequence-specific RNA-binding activity. *Genes & Development*, *3*(10), 1562–1571.

Ner-Gaon, H., Halachmi, R., Savaldi-Goldstein, S., Rubin, E., Ophir, R., & Fluhr, R. (2004). Intron retention is a major phenomenon in alternative splicing in Arabidopsis. *The Plant Journal : for Cell and Molecular Biology*, *39*(6), 877–885. http://doi.org/10.1111/j.1365-313X.2004.02172.x

nheimo, N. I., Gauthier, A., Salojärvi, J., Siligato, R., Brosché, M., Kollist, H., *et al*. (2014). The Arabidopsis thaliana cysteine-rich receptor-like kinases CRK6 and CRK7 protect against apoplastic oxidative stress. *Biochemical and Biophysical Research Communications*, *445*(2), 457–462. http://doi.org/10.1016/j.bbrc.2014.02.013

Nolan, K. E., Kurdyukov, S., & Rose, R. J. (2011). Characterisation of the legume SERK-NIK gene superfamily including splice variants: implications for development and defence. *BMC Plant Biology*, *11*, 44–44. http://doi.org/10.1186/1471-2229-11-44

Nookaew, I., Papini, M., Pornputtapong, N., Scalcinati, G., Fagerberg, L., Uhlén, M., & Nielsen, J. (2012). A comprehensive comparison of RNA-Seq-based transcriptome analysis from reads to differential gene expression and cross-comparison with microarrays: a case study in Saccharomyces cerevisiae. *Nucleic Acids Research*, *40*(20), 10084–10097. http://doi.org/10.1093/nar/gks804

Okada, M., Matsumura, M., Ito, Y., & Shibuya, N. (2002). High-Affinity Binding Proteins for N-Acetylchitooligosaccharide Elicitor in the Plasma Membranes from Wheat, Barley and Carrot Cells: Conserved Presence and Correlation with the Responsiveness to the Elicitor. *Plant and Cell Physiology*, *43*(5), 505–512.

Okada-Katsuhata, Y., Yamashita, A., Kutsuzawa, K., Izumi, N., Hirahara, F., & Ohno, S. (2012). N- and C-terminal Upf1 phosphorylations create binding platforms for SMG-6 and SMG-

5:SMG-7 during NMD. *Nucleic Acids Research*, *40*(3), 1251–1266. http://doi.org/10.1093/nar/gkr791

Palma, K., Zhao, Q., Cheng, Y. T., Bi, D., Monaghan, J., Cheng, W., *et al*. (2007). Regulation of plant innate immunity by three proteins in a complex conserved across the plant and animal kingdoms., *21*(12), 1484–1493. http://doi.org/10.1101/gad.1559607

Palusa, S. G., & Reddy, A. S. N. (2010). Extensive coupling of alternative splicing of pre-mRNAs of serine/arginine (SR) genes with nonsense-mediated decay. *New Phytologist*, *185*(1), 83–89. http://doi.org/10.1111/j.1469-8137.2009.03065.x

Palusa, S. G., Ali, G. S., & Reddy, A. S. N. (2007). Alternative splicing of pre-mRNAs of Arabidopsis serine/arginine-rich proteins: regulation by hormones and stresses. *The Plant Journal : for Cell and Molecular Biology*, *49*(6), 1091–1107. http://doi.org/10.1111/j.1365-313X.2006.03020.x

Pandit, S., Zhou, Y., Shiue, L., Coutinho-Mansfield, G., Li, H., Qiu, J., *et al*. (2013). Genome-wide Analysis Reveals SR Protein Cooperation and Competition in Regulated Splicing. *Molecular Cell*, *50*(2), 223–235. http://doi.org/10.1016/j.molcel.2013.03.001

Parker, J. E., Holub, E. B., Frost, L. N., Falk, A., Gunn, N. D., & Daniels, M. J. (1996). Characterization of eds1, a mutation in Arabidopsis suppressing resistance to Peronospora parasitica specified by several different RPP genes. *Plant Cell*, *8*(11), 2033–2046. http://doi.org/10.1105/tpc.8.11.2033

Patro, R., Mount, S. M., & Kingsford, C. (2014). Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. *Nature Biotechnology*, *32*(5), 462–464. http://doi.org/10.1038/nbt.2862

Pauwels, L., Barbero, G. F., Geerinck, J., Tilleman, S., Grunewald, W., Pérez, A. C., *et al*. (2010). NINJA connects the co-repressor TOPLESS to jasmonate signalling. *Nature*, *464*(7289), 788–791. http://doi.org/10.1038/nature08854

Pertea, M., Mount, S. M., & Salzberg, S. L. (2007). A computational survey of candidate exonic splicing enhancer motifs in the model plant Arabidopsis thaliana. *BMC Bioinformatics*, *8*, 159. http://doi.org/10.1186/1471-2105-8-159

Petersen, M., Brodersen, P., Naested, H., Andreasson, E., Lindhart, U., Johansen, B., *et al*. (2000). Arabidopsis map kinase 4 negatively regulates systemic acquired resistance. *Cell*, *103*(7), 1111–1120. http://doi.org/10.1016/S0092-8674(00)00213-0

Petutschnig, E. K., Jones, A. M. E., Serazetdinova, L., Lipka, U., & Lipka, V. (2010). The lysin motif receptor-like kinase (LysM-RLK) CERK1 is a major chitin-binding protein in Arabidopsis thaliana and subject to chitin-induced phosphorylation. *Journal of Biological Chemistry*, *285*(37), 28902–28911. http://doi.org/10.1074/jbc.M110.116657

Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research*, *29*(9), e45.

Pieterse, C. M. J., Leon-Reyes, A., Van der Ent, S., & Van Wees, S. C. M. (2009). Networking by small-molecule hormones in plant immunity. *Nature Chemical Biology*, *5*(5), 308–316. http://doi.org/10.1038/nchembio.164

Poinssot, B., Vandelle, E., Bentéjac, M., Adrian, M., Levis, C., Brygoo, Y., *et al*. (2003). The endopolygalacturonase 1 from Botrytis cinerea activates grapevine defense reactions unrelated to its enzymatic activity. *Molecular Plant-Microbe Interactions*, *16*(6), 553–564. http://doi.org/10.1094/MPMI.2003.16.6.553

Pré, M., Atallah, M., Champion, A., De Vos, M., Pieterse, C. M. J., & Memelink, J. (2008). The AP2/ERF domain transcription factor ORA59 integrates jasmonic acid and ethylene signals in plant defense. *Plant Physiology*, *147*(3), 1347–1357. http://doi.org/10.1104/pp.108.117523

Qiu, J.-L., Fiil, B. K., Petersen, K., Nielsen, H. B., Botanga, C. J., Thorgrimsen, S., *et al*. (2008). Arabidopsis MAP kinase 4 regulates gene expression through transcription factor release in the nucleus. *EMBO Journal*, *27*(16), 2214–2221. http://doi.org/10.1038/emboj.2008.147

Raj, A., & van Oudenaarden, A. (2008). Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. *Cell*, *135*(2), 216–226. http://doi.org/10.1016/j.cell.2008.09.050

Ramchatesingh, J., Zahler, A. M., Neugebauer, K. M., Roth, M. B., & Cooper, T. A. (1995). A subset of SR proteins activates splicing of the cardiac troponin T alternative exon by direct interactions with an exonic enhancer. *Molecular and Cellular Biology*, *15*(9), 4898–4907.

Rapaport, F., Khanin, R., Liang, Y., Pirun, M., Krek, A., Zumbo, P., *et al*. (2013). Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. *Genome Biology*, *14*(9), R95–R95. http://doi.org/10.1186/gb-2013-14-9-r95

Rayson, S., Arciga-Reyes, L., Wootton, L., De Torres Zabala, M., Truman, W., Graham, N., *et al*. (2012). A role for nonsense-mediated mRNA decay in plants: pathogen responses are induced in Arabidopsis thaliana NMD mutants. *PloS One*, *7*(2), e31917–e31917. http://doi.org/10.1371/journal.pone.0031917

Reddy, A. S. N. (2007). Alternative splicing of pre-messenger RNAs in plants in the genomic era. *Annual Review of Plant Biology*, *58*, 267–294. http://doi.org/10.1146/annurev.arplant.58.032806.103754

Reddy, A. S. N., Day, I. S., Göhring, J., & Barta, A. (2012a). Localization and Dynamics of Nuclear Speckles in Plants. *Plant Physiology*. 158: 67-77. doi:10.1104/pp.111.186700

Reddy, A. S. N., Marquez, Y., Kalyna, M., & Barta, A. (2013). Complexity of the Alternative Splicing Landscape in Plants. *The Plant Cell*. 25(10):3657-83. doi: 10.1105/tpc.113.117523.

Reddy, A. S. N., Rogers, M. F., Richardson, D. N., Hamilton, M., & Ben-Hur, A. (2012b). Deciphering the plant splicing code: experimental and computational approaches for predicting alternative splicing and splicing regulatory elements. *Frontiers in Plant Science*, *3*, 18. http://doi.org/10.3389/fpls.2012.00018

Remy, E., Cabrito, T. R., Batista, R. A., Hussein, M. A. M., Teixeira, M. C., Athanasiadis, A., *et al*. (2014). Intron retention in the 5'UTR of the novel ZIF2 transporter enhances translation to promote zinc tolerance in arabidopsis. *PLoS Genetics*, *10*(5), e1004375. http://doi.org/10.1371/journal.pgen.1004375

Ren, D., Liu, Y., Yang, K.-Y., Han, L., Mao, G., Glazebrook, J., & Zhang, S. (2008). A fungal-responsive MAPK cascade regulates phytoalexin biosynthesis in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(14), 5638–5643. http://doi.org/10.1073/pnas.0711301105

Reyes, T. H., Maekawa, S., Sato, T., & Yamaguchi, J. (2015). The Arabidopsis ubiquitin ligase ATL31 is transcriptionally controlled by WRKY33 transcription factor in response to pathogen attack. *Plant Biotechnology*. 32: 11-19. DOI:10.5511/plantbiotechnology.14.1201b

Richard, H., Schulz, M. H., Sultan, M., Nurnberger, A., Schrinner, S., Balzereit, D., *et al*. (2010). Prediction of alternative isoforms from exon expression levels in RNA-Seq experiments. *Nucleic Acids Research*, *38*(10), e112–e112. http://doi.org/10.1093/nar/gkq041

Richardson, D. N., Rogers, M. F., Labadorf, A., Ben-Hur, A., Guo, H., Paterson, A. H., & Reddy, A. S. N. (2011). Comparative analysis of serine/arginine-rich proteins across 27 eukaryotes: insights into sub-family classification and extent of alternative splicing. *PloS One*, *6*(9), e24542. http://doi.org/10.1371/journal.pone.0024542.s016

Riehs-Kearnan, N. N., Gloggnitzer, J. J., Dekrout, B. B., Jonak, C. C., & Riha, K. K. (2012). Aberrant growth and lethality of Arabidopsis deficient in nonsense-mediated RNA decay factors is caused by autoimmune-like response. *Nucleic Acids Research*, *40*(12), 5615–5624. http://doi.org/10.1093/nar/gks195

Ritchie, M. E., Phipson, B., Di Wu, Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, *43*(7), e47–e47. http://doi.org/10.1093/nar/gkv007

Robinson, M. D., & Smyth, G. K. (2008). Small-sample estimation of negative binomial dispersion, with applications to SAGE data. *Biostatistics*, *9*(2), 321–332. http://doi.org/10.1093/biostatistics/kxm030

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2009). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)*, *26*(1), 139–140. http://doi.org/10.1093/bioinformatics/btp616

Roscigno, R. F., & Garcia-Blanco, M. A. (1995). SR proteins escort the U4/U6.U5 tri-snRNP to the spliceosome. *RNA (New York, N.Y.)*, *1*(7), 692–706.

Roux, M., Schwessinger, B., Albrecht, C., Chinchilla, D., Jones, A., Holton, N., *et al*. (2011). The Arabidopsis leucine-rich repeat receptor-like kinases BAK1/SERK3 and BKK1/SERK4 are required for innate immunity to hemibiotrophic and biotrophic pathogens. *Plant Cell*, *23*(6), 2440–2455. http://doi.org/10.1105/tpc.111.084301

Rushton, P. J., Torres, J. T., Parniske, M., Wernert, P., Hahlbrock, K., & Somssich, I. E. (1996). Interaction of elicitor-induced DNA-binding proteins with elicitor response elements in the promoters of parsley PR1 genes. *EMBO Journal*, *15*(20), 5690–5700.

Ryan, M. C., Cleland, J., Kim, R., Wong, W. C., & Weinstein, J. N. (2012). SpliceSeq: a resource for analysis and visualization of RNA-Seq data on alternative splicing and its functional impacts., *28*(18), 2385–2387. http://doi.org/10.1093/bioinformatics/bts452

Saul, H., Elharrar, E., Gaash, R., Eliaz, D., Valenci, M., Akua, T., *et al*. (2009). The upstream open reading frame of the Arabidopsis AtMHX gene has a strong impact on transcript accumulation through the nonsense-mediated mRNA decay pathway. *The Plant Journal : for Cell and Molecular Biology*, *60*(6), 1031–1042. http://doi.org/10.1111/j.1365-313X.2009.04021.x

Sayani, S., Janis, M., Lee, C. Y., Toesca, I., & Chanfreau, G. F. (2008). Widespread impact of nonsense-mediated mRNA decay on the yeast intronome. *Molecular Cell*, *31*(3), 360–370. http://doi.org/10.1016/j.molcel.2008.07.005

Schöning, J. C., Streitner, C., Meyer, I. M., Gao, Y., & Staiger, D. (2008). Reciprocal regulation of glycine-rich RNA-binding proteins via an interlocked feedback loop coupling alternative splicing to nonsense-mediated decay in Arabidopsis. *Nucleic Acids Research,* 36(22):6977-87. doi: 10.1093/nar/gkn847.

Schöning, J. C., Streitner, C., Page, D. R., Hennig, S., Uchida, K., Wolf, E., *et al*. (2007). Auto-regulation of the circadian slave oscillator component AtGRP7 and regulation of its targets is impaired by a single RNA recognition motif point mutation. *The Plant Journal : for Cell and Molecular Biology*, *52*(6), 1119–1130. http://doi.org/10.1111/j.1365-313X.2007.03302.x

Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., *et al*. (2006). The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Molecular Biology*, *7*, 3. http://doi.org/10.1186/1471-2199-7-3

Schuhegger, R., Nafisi, M., Mansourova, M., Petersen, B. L., Olsen, C. E., Svatos, A., *et al*. (2006). CYP71B15 (PAD3) catalyzes the final step in camalexin biosynthesis. *Plant Physiology*, *141*(4), 1248–1254. http://doi.org/10.1104/pp.106.082024

Seo, P. J., Kim, M. J., Ryu, J.-Y., Jeong, E.-Y., & Park, C.-M. (2011). Two splice variants of the IDD14 transcription factor competitively form nonfunctional heterodimers which may regulate starch metabolism. *Nature Communications*, *2*, 303. http://doi.org/10.1038/ncomms1303

Seok, J., Xu, W., Gao, H., Davis, R. W., & Xiao, W. (2012). JETTA: junction and exon toolkits for transcriptome analysis., *28*(9), 1274–1275. http://doi.org/10.1093/bioinformatics/bts134

Sham, A., Moustafa, K., Al-Ameri, S., Al-Azzawi, A., Iratni, R., & Abuqamar, S. (2015). Identification of Arabidopsis candidate genes in response to biotic and abiotic stresses using comparative microarrays. *PloS One*, *10*(5), e0125666–e0125666. http://doi.org/10.1371/journal.pone.0125666

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., *et al*. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, *13*(11), 2498–2504. http://doi.org/10.1101/gr.1239303

Shen, H., & Green, M. R. (2006). RS domains contact splicing signals and promote splicing by a common mechanism in yeast through humans. *Genes & Development*, *20*(13), 1755–1765. http://doi.org/10.1101/gad.1422106

Sheth, N., Roca, X., Hastings, M. L., Roeder, T., Krainer, A. R., & Sachidanandam, R. (2006). Comprehensive splice-site analysis using comparative genomics. *Nucleic Acids Research*. 34(14):3955-67. DOI:10.1093/nar/gkl556

Shi, C., Baldwin, I. T., & Wu, J. (2012). Arabidopsis Plants Having Defects in Nonsense-mediated mRNA Decay Factors UPF1, UPF2, and UPF3 Show Photoperiod-dependent Phenotypes in Development and Stress Responses. *Journal of Integrative Plant Biology*, *54*(2), 99–114. http://doi.org/10.1111/j.1744-7909.2012.01093.x

Shim, J. S., Jung, C., Lee, S., Min, K., Lee, Y.-W., Choi, Y., *et al*. (2013). AtMYB44 regulates WRKY70 expression and modulates antagonistic interaction between salicylic acid and jasmonic acid signaling. *The Plant Journal : for Cell and Molecular Biology*, *73*(3), 483–495. http://doi.org/10.1111/tpj.12051

Shimizu, T., Nakano, T., Takamizawa, D., Desaki, Y., Ishii-Minami, N., Nishizawa, Y., *et al*. (2010). Two LysM receptor molecules, CEBiP and OsCERK1, cooperatively regulate chitin elicitor signaling in rice. *The Plant Journal*, *64*(2), 204–214. http://doi.org/10.1111/j.1365-313X.2010.04324.x

Shin, C., & Manley, J. L. (2002). The SR protein SRp38 represses splicing in M phase cells. *Cell*, *111*(3), 407–417.

Shinozaki, K., & Sakakibara, H. (2009). Omics and Bioinformatics: An Essential Toolbox for Systems Analyses of Plant Functions Beyond 2010. *Plant and Cell Physiology*, *50*(7), 1177–1180. http://doi.org/10.1093/pcp/pcp085

Shinya, T., Yamaguchi, K., Desaki, Y., Yamada, K., Narisawa, T., Kobayashi, Y., *et al*. (2014). Selective regulation of the chitin-induced defense response by the Arabidopsis receptor-like cytoplasmic kinase PBL27. *The Plant Journal : for Cell and Molecular Biology*, *79*(1), 56–66. http://doi.org/10.1111/tpj.12535

Shiu, S. H. S., & Bleecker, A. B. A. (2001). Receptor-like kinases from Arabidopsis form a monophyletic gene family related to animal receptor kinases. *Pnas*, *98*(19), 10763–10768. http://doi.org/10.1073/pnas.181141598

Shiu, S. H., Karlowski, W. M., Pan, R., Tzeng, Y.-H., Mayer, K. F. X., & Li, W.-H. (2004). Comparative analysis of the receptor-like kinase family in Arabidopsis and rice. *Plant Cell*, *16*(5), 1220–1234. http://doi.org/10.1105/tpc.020834

Simpson, C. G., Fuller, J., Maronova, M., Kalyna, M., Davidson, D., McNicol, J., *et al*. (2008). Monitoring changes in alternative precursor messenger RNA splicing in multiple gene transcripts. *The Plant Journal*, *53*(6), 1035–1048. http://doi.org/10.1111/j.1365-313X.2007.03392.x

Simpson, C. G., Lewandowska, D., Liney, M., Davidson, D., Chapman, S., Fuller, J., *et al*. (2014). Arabidopsis PTB1 and PTB2 proteins negatively regulate splicing of a mini-exon splicing reporter and affect alternative splicing of endogenous genes differentially. *New Phytologist*, *203*(2), 424–436. http://doi.org/10.1111/nph.12821

Simpson, C. G., Manthri, S., Raczynska, K. D., Kalyna, M., Lewandowska, D., Kusenda, B., *et al*. (2010). Regulation of plant gene expression by alternative splicing. *Biochemical Society Transactions*, *38*(2), 667. http://doi.org/10.1042/BST0380667

Song, E. J., Werner, S. L., Neubauer, J., Stegmeier, F., Aspden, J., Rio, D., *et al*. (2010). The Prp19 complex and the Usp4Sart3 deubiquitinating enzyme control reversible ubiquitination at the spliceosome. *Genes & Development*, *24*(13), 1434–1447. http://doi.org/10.1101/gad.1925010

Spadaro, D., Garibaldi, A., & Gullino, M. L. (2004). Control of Penicillium expansum and Botrytis cinerea on apple combining a biocontrol agent with hot water dipping and acibenzolar-S-methyl, baking soda, or ethanol application. *Postharvest Biology and Technology*, *33*(2), 141–151. http://doi.org/10.1016/j.postharvbio.2004.02.002

Staffa, A., & Cochrane, A. (1995). Identification of positive and negative splicing regulatory elements within the terminal tat-rev exon of human immunodeficiency virus type 1. *Molecular and Cellular Biology*, *15*(8), 4597–4605.

Staiger, D., & Apel, K. (1999). Circadian clock-regulated expression of an RNA-binding protein in Arabidopsis: characterisation of a minimal promoter element. *Molecular & General Genetics : MGG*, *261*(4-5), 811–819.

Staiger, D., & Brown, J. W. S. (2013). Alternative Splicing at the Intersection of Biological Timing, Development, and Stress Responses. *The Plant Cell*, *25*(10), 3640–3656. http://doi.org/10.1105/tpc.113.113803

Staiger, D., Zecca, L., Kirk, D. A. W., Apel, K., & Eckstein, L. (2003). The circadian clock regulated RNA-binding protein AtGRP7 autoregulates its expression by influencing alternative splicing of its own pre-mRNA. *The Plant Journal : for Cell and Molecular Biology*, *33*(2), 361–371.

Stamm, S., Ben-Ari, S., Rafalska, I., Tang, Y., Zhang, Z., Toiber, D., *et al*. (2005). Function of alternative splicing. *Gene*, *344*, 1–20. http://doi.org/10.1016/j.gene.2004.10.022

Stauffer, E., Westermann, A., Wagner, G., & Wachter, A. (2010). Polypyrimidine tract-binding protein homologues from Arabidopsis underlie regulatory circuits based on alternative splicing and downstream control. *The Plant Journal*, *64*(2), 243–255. http://doi.org/10.1111/j.1365-313X.2010.04321.x

Steijger, T., Abril, J. F., Engström, P. G., Kokocinski, F., Abril, J. F., Akerman, M., *et al*. (2013). Assessment of transcript reconstruction methods for RNA-seq. *Nature Methods*, *10*(12), 1177–1184. http://doi.org/10.1038/nmeth.2714

Streitner, C., Hennig, L., Korneli, C., & Staiger, D. (2010). Global transcript profiling of transgenic plants constitutively overexpressing the RNA-binding protein AtGRP7. *BMC Plant Biology*, *10*, 221. http://doi.org/10.1186/1471-2229-10-221

Streitner, C., Koster, T., Simpson, C. G., Shaw, P., Danisman, S., Brown, J. W. S., & Staiger, D. (2012). An hnRNP-like RNA-binding protein affects alternative splicing by in vivo interaction with transcripts in Arabidopsis thaliana. *Nucleic Acids Research*, *40*(22), 11240–11255. http://doi.org/10.1093/nar/gks873

Sugnet, C. W., Kent, W. J., Ares, M., & Haussler, D. (2004). Transcriptome and genome conservation of alternative splicing events in humans and mice. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, 66–77.

Supek, F., Bošnjak, M., Škunca, N., & Šmuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. *PloS One*, *6*(7), e21800–e21800. http://doi.org/10.1371/journal.pone.0021800

Syed, N. H., Kalyna, M., Marquez, Y., Barta, A., & Brown, J. W. S. (2012). Alternative splicing in plants--coming of age. *Trends in Plant Science*, *17*(10), 616–623. http://doi.org/10.1016/j.tplants.2012.06.001

Tanabe, N., Yoshimura, K., Kimura, A., Yabuta, Y., & Shigeoka, S. (2007). Differential expression of alternatively spliced mRNAs of Arabidopsis SR protein homologs, atSR30 and atSR45a, in response to environmental stress. *Plant and Cell Physiology*, *48*(7), 1036–1049. http://doi.org/10.1093/pcp/pcm069

Thanaraj, T. A., Stamm, S., Clark, F., Riethoven, J.-J., Le Texier, V., & Muilu, J. (2004). ASD: the Alternative Splicing Database. *Nucleic Acids Research*, *32*(Database issue), D64–9. http://doi.org/10.1093/nar/gkh030

Thines, B., Katsir, L., Melotto, M., Niu, Y., Mandaokar, A., Liu, G., *et al*. (2007). JAZ repressor proteins are targets of the SCFCOI1 complex during jasmonate signalling. *Nature*, *448*(7154), 661–665. http://doi.org/10.1038/nature05960

Thireault, C., Shyu, C., Yoshida, Y., St Aubin, B., Campos, M. L., & Howe, G. A. (2015). Repression of jasmonate signaling by a non-TIFY JAZ protein in Arabidopsis. *The Plant Journal*, *82*(4), 669–679. http://doi.org/10.1111/tpj.12841

Thomas, J., Palusa, S. G., Prasad, K. V. S. K., Ali, G. S., Surabhi, G.-K., Ben-Hur, A., *et al*. (2012). Identification of an intronic splicing regulatory element involved in auto-regulation of alternative splicing of SCL33pre-mRNA. *The Plant Journal*, *72*(6), 935–946. http://doi.org/10.1111/tpj.12004

Thomma, B. P., Eggermont, K., Penninckx, I. A., Mauch-Mani, B., Vogelsang, R., Cammue, B. P., & Broekaert, W. F. (1998). Separate jasmonate-dependent and salicylate-dependent defense-response pathways in Arabidopsis are essential for resistance to distinct microbial pathogens. *Pnas*, *95*(25), 15107–15111.

Trapnell, C. C., Pachter, L. L., & Salzberg, S. L. S. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics (Oxford, England)*, *25*(9), 1105–1111. http://doi.org/10.1093/bioinformatics/btp120

Trapnell, C., Hendrickson, D. G., Sauvageau, M., Goff, L., Rinn, J. L., & Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature Biotechnology*, *31*(1), 46–53. http://doi.org/10.1038/nbt.2450

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., *et al*. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology*, *28*(5), 511–515. http://doi.org/10.1038/nbt.1621

Tsuda, K., & Somssich, I. E. (2015). Transcriptional networks in plant immunity. *New Phytologist*, *206*(3), 932–947. http://doi.org/10.1111/nph.13286

Tsuda, K., Sato, M., Stoddard, T., Glazebrook, J., & Katagiri, F. (2009). Network properties of robust immunity in plants. *PLoS Genetics*, *5*(12), e1000772. http://doi.org/10.1371/journal.pgen.1000772

Tzafrir, I., Dickerman, A., Brazhnik, O., Nguyen, Q., McElver, J., Frye, C., *et al*. (2003). The Arabidopsis SeedGenes Project. *Nucleic Acids Research,* 31(1): 90–93.

Ule, J., Stefani, G., Mele, A., Ruggiu, M., Wang, X., Taneri, B., *et al*. (2006). An RNA map predicting Nova-dependent splicing regulation. *Nature*, *444*(7119), 580–586. http://doi.org/10.1038/nature05304

Vailleau, F., Daniel, X., Tronchet, M., Montillet, J.-L., Triantaphylidès, C., & Roby, D. (2002). A R2R3-MYB gene, AtMYB30, acts as a positive regulator of the hypersensitive cell death program in plants in response to pathogen attack. *Pnas*, *99*(15), 10179–10184. http://doi.org/10.1073/pnas.152047199

Valenzuela, C. E., Acevedo-Acevedo, O., Miranda, G. S., Vergara-Barros, P., Holuigue, L., Figueroa, C. R., & Figueroa, P. M. (2016). Salt stress response triggers activation of the jasmonate signaling pathway leading to inhibition of cell elongation in Arabidopsis primary root. *Journal of Experimental Botany*. 67 (14), 4209-4220 http://doi.org/10.1093/jxb/erw202

Van der Does, D., Leon-Reyes, A., Koornneef, A., Van Verk, M. C., Rodenburg, N., Pauwels, L., *et al*. (2013). Salicylic acid suppresses jasmonic acid signaling downstream of SCFCOI1-JAZ by targeting GCC promoter motifs via transcription factor ORA59. *Plant Cell*, *25*(2), 744–761. http://doi.org/10.1105/tpc.112.108548

Wachter, A. (2014). Gene regulation by structured mRNA elements. *trends in Genetics*, 30 (5), 172-181

Wachter, A., Rühl, C., & Stauffer, E. (2012). The Role of Polypyrimidine Tract-Binding Proteins and Other hnRNP Proteins in Plant Splicing Regulation. *Frontiers in Plant Science*, *3*, 81–81. http://doi.org/10.3389/fpls.2012.00081

Wager, A., & Browse, J. (2012). Social Network: JAZ Protein Interactions Expand Our Knowledge of Jasmonate Signaling. *Frontiers in Plant Science*, *3*, 41–41. http://doi.org/10.3389/fpls.2012.00041

Wagner, G. P., Kin, K., & Lynch, V. J. (2012). Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory in Biosciences*, *131*(4), 281–285. http://doi.org/10.1007/s12064-012-0162-3

Wahl, M. C., Will, C. L., & Lührmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. *Cell*, *136*(4), 701–718. http://doi.org/10.1016/j.cell.2009.02.009

Wan, J., Tanaka, K., Zhang, X.-C., Son, G. H., Brechenmacher, L., Nguyen, T. H. N., & Stacey, G. (2012). LYK4, a lysin motif receptor-like kinase, is important for chitin signaling and plant innate immunity in Arabidopsis. *Plant Physiology*, *160*(1), 396–406. http://doi.org/10.1104/pp.112.201699

Wan, J., Zhang, X.-C., Neece, D., Ramonell, K. M., Clough, S., Kim, S.-Y., *et al*. (2008). A LysM receptor-like kinase plays a critical role in chitin signaling and fungal resistance in Arabidopsis. *Plant Cell*, *20*(2), 471–481. http://doi.org/10.1105/tpc.107.056754

Wang, B.-B., & Brendel, V. (2004a). Genomewide comparative analysis of alternative splicing in plants. *Pnas*, *103*, 7175–7180.

Wang, B.-B., & Brendel, V. (2004b). The ASRG database: identification and survey of Arabidopsis thaliana genes involved in pre-mRNA splicing. *Genome Biol*, *5*(12), R102. http://doi.org/10.1186/gb-2004-5-12-r102

Wang, E. T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., *et al*. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature*, *456*(7221), 470–476. http://doi.org/10.1038/nature07509

Wang, K., Singh, D., Zeng, Z., Coleman, S. J., Huang, Y., Savich, G. L., *et al*. (2010). MapSplice: Accurate mapping of RNA-seq reads for splice junction discovery. *Nucleic Acids Research*, *38*(18), e178–e178. http://doi.org/10.1093/nar/gkq622

Wang, Y., Xiao, X., Zhang, J., Choudhury, R., Robertson, A., Li, K., *et al*. (2012). A complex network of factors with overlapping affinities represses splicing through intronic elements. *Nature Structural & Molecular Biology*, *20*(1), 36–45. http://doi.org/10.1038/nsmb.2459

Williamson, B., Tudzynski, B., Tudzynski, P., & van Kan, J. A. L. (2007). Botrytis cinerea: the cause of grey mould disease. *Molecular Plant Pathology*, *8*(5), 561–580. http://doi.org/10.1111/j.1364-3703.2007.00417.x

Windram, O., Madhou, P., McHattie, S., Hill, C., Hickman, R., Cooke, E., *et al*. (2012). Arabidopsis Defense against Botrytis cinerea: Chronology and Regulation Deciphered by High-Resolution Temporal Transcriptomic Analysis. *Plant Cell*, *24*(9), 3530–3557. http://doi.org/10.1105/tpc.112.102046

Windram, O., Stoker, C., & Denby, K. (2015). Overview of Plant Defence Systems: Lessons from Arabidopsis-Botrytis cinerea Systems Biology (pp. 335–360). Cham: Springer International Publishing. http://doi.org/10.1007/978-3-319-23371-0_17

Woloshen, V. V., Huang, S. S., & Li, X. X. (2011). RNA-Binding Proteins in Plant Immunity. *Journal of Pathogens*, *2011*, 278697–278697. http://doi.org/10.4061/2011/278697

Wu, J. Y., & Maniatis, T. (1993). Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell*, *75*(6), 1061–1070. http://doi.org/10.1016/0092-8674(93)90316-I

Wu, J., Anczuków, O., Krainer, A. R., Zhang, M. Q., & Zhang, C. (2013). OLego: fast and sensitive mapping of spliced mRNA-Seq reads using small seeds. *Nucleic Acids Research*, *41*(10), 5149–5163. http://doi.org/10.1093/nar/gkt216

Wu, L.-T., Zhong, G.-M., Wang, J.-M., Li, X.-F., Song, X., & Yang, L. (2011). Arabidopsis WRKY28 transcription factor is required for resistance to necrotrophic pathogen, Botrytis cinerea. *African Journal of Microbiology Research*, *5*(30), 5481–5488. http://doi.org/10.5897/AJMR11.781

Wu, T. D., & Nacu, S. (2010). Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics (Oxford, England)*, *26*(7), 873–881. http://doi.org/10.1093/bioinformatics/btq057

Xie, D.-X., Feys, B. F., James, S., Nieto-Rostro, M., & Turner, J. G. (1998). COI1: An Arabidopsis Gene Required for Jasmonate-Regulated Defense and Fertility. *Science*, *280*, 1091–1094.

Xu, F., Xu, S., Wiermer, M., Zhang, Y., & Li, X. (2012). The cyclin L homolog MOS12 and the MOS4-associated complex are required for the proper splicing of plant resistance genes. *The Plant Journal*, *70*(6), 916–928. http://doi.org/10.1111/j.1365-313X.2012.04906.x

Xu, J., Meng, J., Meng, X., Zhao, Y., Liu, J., Sun, T., *et al*. (2016). Pathogen-responsive MPK3 and MPK6 reprogram the biosynthesis of indole glucosinolates and their derivatives in Arabidopsis immunity. *The Plant Cell*. 28(5), 1144-62. http://doi.org/10.1105/tpc.15.00871

Xu, L. (2002). The SCFCOI1 Ubiquitin-Ligase Complexes Are Required for Jasmonate Response in Arabidopsis. *The Plant Cell Online*, *14*(8), 1919–1935. http://doi.org/10.1105/tpc.003368

Xu, S., Zhang, Z., Jing, B., Gannon, P., Ding, J., Xu, F., *et al*. (2011). Transportin-SR is required for proper splicing of resistance genes and plant immunity. *PLoS Genetics*, *7*(6), e1002159–e1002159. http://doi.org/10.1371/journal.pgen.1002159

Xu, X., Chen, C., Fan, B., & Chen, Z. (2006). Physical and functional interactions between pathogen-induced Arabidopsis WRKY18, WRKY40, and WRKY60 transcription factors. *Plant Cell*, *18*(5), 1310–1326. http://doi.org/10.1105/tpc.105.037523

Xu, Y.-Z., Newnham, C. M., Kameoka, S., Huang, T., Konarska, M. M., & Query, C. C. (2004). Prp5 bridges U1 and U2 snRNPs and enables stable U2 snRNP association with intron RNA. *EMBO Journal*, *23*(2), 376–385. http://doi.org/10.1038/sj.emboj.7600050

Yamaguchi, K., Yamada, K., Ishikawa, K., Yoshimura, S., Hayashi, N., Uchihashi, K., *et al*. (2013). A receptor-like cytoplasmic kinase targeted by a plant pathogen effector is directly phosphorylated by the chitin receptor and mediates rice immunity. *Cell Host & Microbe*, *13*(3), 347–357. http://doi.org/10.1016/j.chom.2013.02.007

Yamashita, A., Kashima, I., & Ohno, S. (2005). The role of SMG-1 in nonsense-mediated mRNA decay. *Biochimica Et Biophysica Acta*, *1754*(1-2), 305–315. http://doi.org/10.1016/j.bbapap.2005.10.002

Yan, Y., Stolz, S., Chetelat, A., Reymond, P., Pagni, M., Dubugnon, L., & Farmer, E. E. (2007). A Downstream Mediator in the Growth Repression Limb of the Jasmonate Pathway. *The Plant Cell*, *19*(8), 2470–2483. http://doi.org/10.1105/tpc.107.050708

Yeakley, J. M., Morfin, J. P., Rosenfeld, M. G., & Fu, X. D. (1996). A complex of nuclear proteins mediates SR protein binding to a purine-rich splicing enhancer. *Pnas*, *93*(15), 7582–7587.

Yoine, M., Ohto, M.-A., Onai, K., Mita, S., & Nakamura, K. (2006). The lba1 mutation of UPF1 RNA helicase involved in nonsense-mediated mRNA decay causes pleiotropic phenotypic changes and altered sugar signalling in Arabidopsis. *The Plant Journal : for Cell and Molecular Biology*, *47*(1), 49–62. http://doi.org/10.1111/j.1365-313X.2006.02771.x

Yoshimura, K., Yabuta, Y., Ishikawa, T., & Shigeoka, S. (2002). Identification of a cis element for tissue-specific alternative splicing of chloroplast ascorbate peroxidase pre-mRNA in higher plants. *Journal of Biological Chemistry*, *277*(43), 40623–40632. http://doi.org/10.1074/jbc.M201531200

Zamore, P. D., & Green, M. R. (1989). Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. *Pnas*, *86*(23), 9243–9247.

Zappia, L., Grillet, F., Molck, C., Pham, K., Pannequin, J., Taylor, G. R., *et al*. (2015). Combining multiple methods to improve differential expression testing [v1; not peer reviewed]. *F1000Research 2015, 4:1015 (Poster)*, 1–1. doi: 10.7490/f1000research.1110759.1

Zarei, A., Körbes, A. P., Younessi, P., Montiel, G., Champion, A., & Memelink, J. (2011). Two GCC boxes and AP2/ERF-domain transcription factor ORA59 in jasmonate/ethylene-mediated activation of the PDF1.2 promoter in Arabidopsis. *Plant Molecular Biology*, *75*(4-5), 321–331. http://doi.org/10.1007/s11103-010-9728-y

Zhang, J., Li, W., Xiang, T., Liu, Z., Laluk, K., Ding, X., *et al*. (2010a). Receptor-like Cytoplasmic Kinases Integrate Signaling from Multiple Plant Immune Receptors and Are Targeted by a

Pseudomonas syringae Effector. *Cell Host & Microbe*, *7*(4), 290–301. http://doi.org/10.1016/j.chom.2010.03.007

Zhang, J., Lu, H., Li, X., Li, Y., Cui, H., Wen, C.-K., *et al*. (2010b). Effector-triggered and pathogen-associated molecular pattern-triggered immunity differentially contribute to basal resistance to Pseudomonas syringae. *Molecular Plant-Microbe Interactions*, *23*(7), 940–948. http://doi.org/10.1094/MPMI-23-7-0940

Zhang, L., Kars, I., Essenstam, B., Liebrand, T. W. H., Wagemakers, L., Elberse, J., *et al*. (2014a). Fungal endopolygalacturonases are recognized as microbe-associated molecular patterns by the arabidopsis receptor-like protein RESPONSIVENESS TO BOTRYTIS POLYGALACTURONASES1. *Plant Physiology*, *164*(1), 352–364. http://doi.org/10.1104/pp.113.230698

Zhang, R., Calixto, C. P. G., Marquez, Y., Venhuizen, P., Tzioutziou, N. A., Guo, W., *et al*. (2016). AtRTD2: A Reference Transcript Dataset for accurate quantification of alternative splicing and expression changes in Arabidopsis thaliana RNA-seq data. *bioRxiv* https://doi.org/10.1101/051938

Zhang, R., Calixto, C. P. G., Tzioutziou, N. A., James, A. B., Simpson, C. G., Guo, W., *et al*. (2015). AtRTD - a comprehensive reference transcript dataset resource for accurate quantification of transcript-specific expression in Arabidopsis thaliana. *New Phytologist*, *208*(1), 96–101. http://doi.org/10.1111/nph.13545

Zhang, X.-C., & Gassmann, W. (2003). RPS4-Mediated Disease Resistance Requires the Combined Presence of RPS4 Transcripts with Full-Length and Truncated Open Reading Frames. *The Plant Cell Online*, *15*(10), 2333–2342.

Zhang, X.-C., & Gassmann, W. (2007). Alternative splicing and mRNA levels of the disease resistance gene RPS4 are induced during defense responses. *Plant Physiology*, *145*(4), 1577–1587. http://doi.org/10.1104/pp.107.108720

Zhang, Y., Goritschnig, S., Dong, X., & Li, X. (2003). A gain-of-function mutation in a plant disease resistance gene leads to constitutive activation of downstream signal transduction pathways in suppressor of npr1-1, constitutive 1. *Plant Cell*, *15*(11), 2636–2646. http://doi.org/10.1105/tpc.015842

Zhang, Z., Liu, Y., Ding, P., Li, Y., Kong, Q., & Zhang, Y. (2014b). Splicing of receptor-like kinase-encoding SNC4 and CERK1 is regulated by two conserved splicing factors that are required for plant immunity. *Molecular Plant*, *7*(12), 1766–1775. http://doi.org/10.1093/mp/ssu103

Zhang, Z., Wu, Y., Gao, M., Zhang, J., Kong, Q., Liu, Y., *et al*. (2012). Disruption of PAMP-induced MAP kinase cascade by a Pseudomonas syringae effector activates plant immunity mediated by the NB-LRR protein SUMM2. *Cell Host & Microbe*, *11*(3), 253–263. http://doi.org/10.1016/j.chom.2012.01.015

Zheng, Z., Qamar, S. A., Chen, Z., & Mengiste, T. (2006). Arabidopsis WRKY33 transcription factor is required for resistance to necrotrophic fungal pathogens. *The Plant Journal : for Cell and Molecular Biology*, *48*(4), 592–605. http://doi.org/10.1111/j.1365-313X.2006.02901.x

Zhu, Z., An, F., Feng, Y., Li, P., Xue, L., A, M., *et al*. (2011). Derepression of ethylene-stabilized transcription factors (EIN3/EIL1) mediates jasmonate and ethylene signaling synergy in Arabidopsis. *Pnas*, *108*(30), 12539–12544. http://doi.org/10.1073/pnas.1103959108

Zipfel, C. (2014). Plant pattern-recognition receptors. *Trends in Immunology*, *35*(7), 345–351. http://doi.org/10.1016/j.it.2014.05.004

7 Appendix

7.1 DESeq2 example code

```
>library("DESeq2")
> rawdata<-
read.delim("~/Dropbox/MAC_KO_mos4_SF_quant_estNoReads.txt",
check.names=FALSE, header=TRUE)
> head(rawdata)
               M1  M2  M3  I1  I2  I3
AT1G01010.1   153 110 202 350 245 139
AT1G01020.1     9   0   0   9   3   0
AT1G01020.2    17  24   6 163  36  67
AT1G01020_ID1  57  42  52  78  58  49
AT1G01020_ID3 144  56 108  14  30  15
AT1G01020_ID4 127 121 151  42 104  62
> countData<-rawdata[,c(4,5,6,1,2,3)]
> head(countData)
               I1  I2  I3  M1  M2  M3
AT1G01010.1   350 245 139 153 110 202
AT1G01020.1     9   3   0   9   0   0
AT1G01020.2   163  36  67  17  24   6
AT1G01020_ID1  78  58  49  57  42  52
AT1G01020_ID3  14  30  15 144  56 108
AT1G01020_ID4  42 104  62 127 121 151
> colData<-read.delim("R_targets_DESeq.txt", check.names=FALSE,
header=TRUE)
> colData$Condition<-
factor(colData$Condition,levels=c("untreated","treated"))
> dds<-
DESeqDataSetFromMatrix(countData=countData,colData=colData,design=~C
ondition)
> dds
```

```
class: DESeqDataSet
dim: 74216 6
metadata(0):
assays(1): counts
rownames(74216): AT1G01010.1 AT1G01020.1 ... ATMG01400.1 ATMG01410.1
rowRanges metadata column names(0):
colnames(6): I1 I2 ... M2 M3
colData names(1): Condition
> dds<-DESeq(dds)
estimating size factors
estimating dispersions
gene-wise dispersion estimates
mean-dispersion relationship
final dispersion estimates
fitting model and testing
> res<-results(dds)
> res<-res[order(res$padj),]
> summary(res)

out of 54951 with nonzero total read count
adjusted p-value < 0.1
LFC > 0 (up)     : 8673, 16%
LFC < 0 (down)   : 7307, 13%
outliers [1]     : 2754, 5%
low counts [2]   : 6250, 11%
(mean count < 2)
[1] see 'cooksCutoff' argument of ?results
[2] see 'independentFiltering' argument of ?results

> resMLE<-results(dds, addMLE=TRUE)
> mcols(res)$description
```

```
[1] "mean of normalized counts for all samples"              "log2
fold change (MAP): Condition treated vs untreated"
[3] "standard error: Condition treated vs untreated"         "Wald
statistic: Condition treated vs untreated"
[5] "Wald test p-value: Condition treated vs untreated"      "BH
adjusted p-values"
> head(results(dds, addMLE=TRUE))
log2 fold change (MAP): Condition treated vs untreated
Wald test p-value: Condition treated vs untreated
DataFrame with 6 rows and 7 columns
               baseMean log2FoldChange     lfcMLE     lfcSE
stat      pvalue        padj
              <numeric>      <numeric>  <numeric> <numeric>
<numeric>   <numeric>   <numeric>
AT1G01010.1   199.43141      1.1038776  1.1229056 0.4286133
2.5754629 0.010010601 0.036542232
AT1G01020.1     3.45435      0.3910145  0.7903599 1.6601116
0.2355351 0.813793448 0.907755762
AT1G01020.2    55.78323      2.6099653  2.8072584 0.8459532
3.0852361 0.002033906 0.009418653
AT1G01020_ID1  55.43492      0.7186443  0.7298446 0.4080630
1.7611112 0.078219592 0.188184920
AT1G01020_ID3  53.06053     -1.8316683 -1.9012880 0.6229000 -
2.9405497 0.003276305 0.014176135
AT1G01020_ID4  94.71296     -0.4630737 -0.4741422 0.5011860 -
0.9239559 0.355509283 0.554673675
> resSig<-subset(res,padj<0.1)
> resAll<-as.data.frame(res)
> MACKO_mos4_Dundee_SF_DESeq2_all<-resAll
> MACKO_mos4_Dundee_SF_DESeq2_Sig<-resSig
>write.table(MACKO_mos4_Dundee_SF_DESeq2_all,"MACKO_mos4_Dundee_SF_D
ESeq2_all.txt",row.name=T,quote=FALSE)
```

```
>write.table(MACKO_mos4_Dundee_SF_DESeq2_Sig,"MACKO_mos4_Dundee_SF_D
ESeq2_Sig.txt",row.name=T,quote=FALSE)
```

7.2 EdgeR example code

```
>library(edgeR)
>target<-read.delim("targets_edgeR")
>rawdata<-read.delim("~/Dropbox/Col0_Tophat_JunctionCount.txt",
header=T)
>y<-DGEList(counts=rawdata[,2:7],group=target$Treatment
genes=rawdata[,1])
>keep<-rowSums(cpm(y)>1)>=3
#filters out v low expressed transcripts that have <1cpm in less
than 3 samples
>y<-y[keep,]
>y$samples$lib.size<-colSums(y$counts)
>y<-calcNormFactors(y)
>y<-estimateCommonDisp(y,verbose=TRUE)
>y<-estimateTagwiseDisp(y)
>et<-exactTest(y, pair=c("M","I")
>top<-topTags(et)
>summary(de<-decideTestsDGE(et))
     [,1]
-1    373
0     16417
1     1428
>top_all<-topTags(et, n=18218)
>Col0_all_edgeR<- top_all
>top_DE<-topTags(et, n=1801)
>Col0_DE_edgeR<- top_DE
>
write.table(Col0_all_edgeR,"Col0_all_edgeR",row.name=T,quote=FALSE)
316
```

```
> write.table(Col0_DE_edgeR,"Col0_DE_edgeR",row.name=T,quote=FALSE)
```

7.3 Limma example code

```
> SJraw<-read.delim("~/Dropbox/Col0_Tophat_JunctionCount.txt",
header=T)
>SJCounts<-SJraw[,2:7]
>SJgenes<-SJraw[,c(1,8,9)]
>SJy<-DGEList(counts=SJcounts, genes=SJgenes)
>isexpr<-rowSums(cpm(SJy)>1)>=3
#filters out SJ that have <1cpm in less than 3 samples
>SJy<-SJy[isexpr, ,keep.lib.sizes.TRUE]
>SJy<-calcNormFactors(SJy)
>targets<-read.delim("Targets.txt")
>Exon_Targets<-factor(targets$Treatment, levels=c("M","I"))
>design<-model.matrix(~Exon_Targets)
>SJvwts<-
voomWithQualityWeights(SJy,design=design,normilization=none,plot=TRU
E)
>vfitSJ<-lmFit(vwtsSJ)
>VfitSJ.de<-eBayes(vfitSJ, robust=TRUE)
>topTable(vfitSJ.de, coef="Exon_TargetsI"
>summary(decideTests(vfitSJ.de)
     Intercept  Exon_TargetsI
-1   511        3874
0    10120      56200
1    57380      7977
>DESJ_TH_limma<-topTable(vfitSJ.de,coef="Exon_TargetsI",
number=11851)
>AllSJ_TH_limma<-topTable(vfitSJ.de,coef="Exon_TargetsI",
number=68051)
> write.table(DESJ_TH_limma,"DESJ_TH_limma",row.name=T,quote=FALSE)
```

```
>
write.table(AllSJ_TH_limma,"AllSJ_TH_limma",row.name=T,quote=FALSE)
```